



**QUEEN'S  
UNIVERSITY  
BELFAST**

## **Malaria Suitability, Urbanization and Persistence: Evidence From China Over More Than 2000 Years**

Flückiger, M., & Ludwig, M. (2017). Malaria Suitability, Urbanization and Persistence: Evidence From China Over More Than 2000 Years. *European Economic Review*, 92, 146–160.  
<https://doi.org/10.1016/j.euroecorev.2016.12.006>

**Published in:**  
European Economic Review

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

### **Publisher rights**

© 2016 Elsevier BV. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>, which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

### **General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

# Malaria Suitability, Urbanization and Persistence: Evidence From China Over More Than 2000 Years\*

Matthias Flückiger<sup>†</sup>

Markus Ludwig<sup>‡</sup>

December 17, 2016

## Abstract

We show that the climatic potential for *Plasmodium falciparum* malaria transmission constituted a locational fundamental that influenced the spatial distribution of urbanization since the early start of the southward expansion of the Han Chinese around 200 BCE. This effect is still detectable in today's distribution of urbanization and economic activity even though the risk of malaria *falciparum* has been successfully eliminated. We do not find any indication of convergence between high- and low malaria potential regions after eradication. Our identification strategy relies on a climate-based measure of *Plasmodium falciparum* malaria transmission intensity which is fitted to experimental data on mosquito and parasite development from laboratory studies. This measure is exogenous with respect to human population densities.

JEL CLASSIFICATION: I15, R12, O18, N15

KEYWORDS: Malaria, Urbanization, Economic Growth, China

---

\*We thank two anonymous referees for comments that substantially improved the paper. We also thank Philipp vom Berge, Anton Bondarev, Frank Krysiak, Hermione Miller-Moser, Jürg H. Sommer, and participants at the 2016 Annual Conference of the Canadian Economics Association for their helpful comments.

<sup>†</sup>Queen's University Belfast and University of Basel. Email: [matthias.flueckiger@qub.ac.uk](mailto:matthias.flueckiger@qub.ac.uk)

<sup>‡</sup>University of Bayreuth and University of Basel. Email: [markus.ludwig@uni-bayreuth.de](mailto:markus.ludwig@uni-bayreuth.de)

# 1 Introduction

In this paper, we analyze the role of *Plasmodium malaria falciparum*—one of history’s leading causes of death and disease (Webb, 2008)—in shaping the spatial distribution of economic activity in South China. The case of China offers a unique opportunity to analyze the role of the climatic malaria environment over a very long time period. The availability of both historical and current-day data on the location of urban centers—a proxy for economic development (Chang (1961, 1963), de Vries (1976, p.164), Paul Bairoch (1988, Chapter 1), Lucas (1988))—allows us to track the effect of the local malaria transmission potential on urbanization patterns of the Han Chinese,<sup>1</sup> today’s dominant population group, over a period of more than 2,000 years from their initial colonization of the malarious regions of South China around 200 BCE up until 2010 CE. An important feature of China’s recent history is the implementation of large-scale eradication campaigns in the 1950s which successfully eliminated the malaria falciparum risk. This implies that we can analyze to what extent malaria-related regional differences in urbanization and economic activity persist after eradication. An additional advantage of focusing on the case of China is that by concentrating on a single country we ensure that the analysis is not affected by potentially confounding differences in country-level characteristics, such as institutions. Increasing the relevance of our study is further the fact that China has by far the largest urban population worldwide.<sup>2</sup>

Our empirical analysis focuses on South China, i.e., the part of today’s China that lies south of the Huai River–Qin Mountains line and in which the majority of China’s population lives. Prior to eradication, malaria falciparum was, in contrast to North China, endemic in this region. We make use of various data sources that provide information regarding the location of urban and political centers at various points in time. This allows us to track the spatial distribution of urbanization from the start of the southward expansion of the Han Chinese around 200 BCE up to 2010. The sources used include CHGIS (2010), Skinner et al. (2011), Skinner et al. (2013) and Brinkhoff (2015). To obtain a time-consistent unit of observation, we aggregate the information contained in these data sources at the 0.5 degree  $\times$  0.5 degree grid-cell level. South China spans over a total of 868 grid cells.

To identify the effect of the malaria environment on economic development, we rely on a

---

<sup>1</sup>Throughout the paper, we will use the terms ‘Han Chinese’ and ‘Chinese’ interchangeably. In both cases, we refer to the people of Han ethnicity.

<sup>2</sup>As of 2010, more than 600 million people live in urban areas. This is almost twice as much as in India, the country with the second largest number of urban dwellers (World Development Indicators).

measure that describes the development of the mosquito as well as the *Plasmodium falciparum* parasite as non-monotonic functions of air temperature. The functional form of the model is derived exclusively from experimental data gained in laboratory studies and is therefore arguably exogenous to the spatial distribution of human population.<sup>3</sup> Our measure can be interpreted as a locational fundamental, i.e., a first nature characteristic. It reflects the local, time-invariant climatic suitability for malaria *falciparum* transmission which not only influences outcome variables via contemporaneous effects on health, but also through its effects on long-term determinants of economic development. For example, a continuously adverse health environment can influence the evolution of local institutions or social and cultural traits. Furthermore, it can deter the establishment of settlements which can, through the existence of increasing returns to scale, influence a region's path of urbanization and economic development. When interpreting our empirical results, it is important to keep in mind that our estimates capture reduced-form effects of the malaria environment. They represent the net effect of all channels through which the malaria environment influences outcome variables. Our cross-sectional OLS regression results document that the climatic potential for malaria transmission constituted a locational fundamental during the Imperial Era (221 BCE–1911 CE), i.e., the pre-eradication era. From the beginning of the Han Chinese colonization of the South—that is, the earliest point in time for which urbanization data for South China are available—we observe that the local variation in the climatic malaria potential strongly influenced the spatial pattern of urbanization. After the successful eradication of malaria *falciparum* that started in the 1950s, the climatic potential for its transmission ceased to pose a threat to human health. Nevertheless, we find that the negative relationship between the local potential for malaria transmission and urbanization as well as economic development persists to the present day. We do not observe any convergence between high- and low-malaria potential areas as a result of eradication-related individual health benefits. The strong degree of persistence in the spatial pattern of urbanization suggests that the negative effects of the malaria environment on long-term determinants of development dominate the positive health-related effects that result from eradication. However, given the short period of 50 years that has passed since the beginning of the malaria eradication campaigns, there is also the possibility that convergence between previously highly malarious areas and less affected regions

---

<sup>3</sup>Note that prevalence-based malaria intensity measures are not exogenous with respect to the spatial distribution of human population. Reverse causality is inherent in the relationship between human population density and malaria prevalence. In a region where humans are absent, the prevalence is zero. Likewise, for malaria to be prevalent among humans, we need the presence of a sufficiently high human population density (see Section 4.1).

has started but is not detectable yet due to a slow pace of the adaptation processes.

The remainder of this paper is organized as follows: In the next section, we discuss the related literature. Section 3 describes the malaria environment in South China and the problems it posed for the Han Chinese during their colonization of the South. Furthermore, we briefly discuss why the grid-cell level number of administrative capitals can be used as a proxy for the degree of urbanization during the imperial era. Next, we outline the properties of our malaria suitability measure and present our empirical strategy. In Section 5, we describe the data employed in the empirical analysis and conduct a descriptive analysis. In Section 6, we address our research question more formally in a regression setup. Finally, Section 7 concludes.

## 2 Related Literature

Our paper is related to the large body of research that analyzes the effect of malaria on aggregate economic development (e.g., Gallup et al. (1999), Acemoglu et al. (2001), Gallup and Sachs (2001), Easterly and Levine (2012) or Depetris-Chauvin and Weil (2013)). These studies are usually conducted at the country level and do not employ an exogenous measure of malaria suitability.<sup>4</sup> The findings are inconclusive, even though mostly suggesting a negative effect of malaria. Compared to the country-level studies, we abstract from cross-country differences in the quality of institutions and document that the local disease environment induces regional inequalities within a given set of national institutions. In that respect, our study relates to our recent paper (Flückiger and Ludwig (2016)) in which we analyze the effect of the malaria environment on the geographical distribution of urbanization in Sub-Saharan Africa. We find that an increase in malaria suitability reduces urbanization as well as economic activity at the sub-national level. Compared to China, however, malaria is still endemic in large parts of Africa. Further differentiating the two studies is the economic system as well as the time span within which the empirical analysis takes place.

Our paper is also closely related to the macroeconomic literature on health and development (Young, 2005; Cervellati and Sunde, 2013; Aghion et al., 2011; Soares, 2005).<sup>5</sup> The theoretical models of Gollin and Zimmermann (2008) as well as Chakraborty et al. (2010)

---

<sup>4</sup>The malaria measures used in the cited papers are usually based on malaria prevalence (see Section 4.1).

<sup>5</sup>See Weil (2014) for a complete review.

specifically analyze the effect of malaria within a general equilibrium framework. Their models suggest that malaria eradication generate large economic gains and lead to convergence between previously highly malarious areas and less affected regions. Contrasting these findings are the results of Acemoglu and Johnson (2007) and Ashraf et al. (2009). They show that a decline in (child) mortality and morbidity does not necessarily increase the level of economic development in the short run. Acemoglu and Johnson (2007) argue that disease eradication improves health but also leads to population growth. When land supply is inelastic, an increase in population reduces output per capita due to the existence of diminishing marginal returns. While our findings are more in line with the latter studies, our analysis suggests that, in the case of China, persistence is not generated by a disproportionate post-eradication population growth in previously highly malarious areas that dilutes health-related economic gains. One possible explanation for the absence of this effect is that our study does not look at general equilibrium effects across countries but within one single nation. Barriers to equalization of marginal returns—the main driver of the results of Acemoglu and Johnson (2007), may be lower within than across countries. An additional explanation is the introduction of the one-child policy that influenced population growth dynamics in China. This implies that persistence results from a variety of complex, malaria-induced differences in long-term determinants of growth. An important aspect in which our paper differs from the studies mentioned above is the type of measure used to capture malaria intensity. Our measure represents a local time-invariant geographic characteristics that influences contemporaneous development through a variety of channels, including its effect on contemporaneous health as well as long-term determinants of growth. Eradication, therefore only affects one channel—the malaria environment’s effects on contemporaneous malaria prevalence—through which our measure influences outcomes. Compared to indirect effects of the disease environment, direct effects on health may be relatively less important in explaining regional variation in development (Acemoglu et al., 2003). By documenting that the malaria environment represents a locational fundamental that has affected local economic development since ancient times, our work also speaks to the literature that investigates the importance of deep, fundamental and historical origins of current-day disparities in economic development.<sup>6</sup> Particularly closely related to our study are papers that document that geographic characteristics, for example disease environments, influence today’s regional development via a variety of complex and interleaved channels, such as institutions or cultural traits, that influence the course of history and

---

<sup>6</sup>See Nunn (2009) or Spolaore and Wacziarg (2013) for an overview of this literature.

economic development (e.g., Acemoglu et al. (2001); Alsan (2015); Nunn and Puga (2012)). Our results further contribute to the discussion regarding the determinants of the spatial distribution of economic activity. There are several papers documenting that natural advantages play a key role in explaining this distribution (e.g., Davis and Weinstein (2002); Rappaport and Sachs (2003); Bosker et al. (2007); Miguel and Roland (2011); Motamed et al. (2014)). Our work also relates to the studies that show that now obsolete natural or man-made factors that represented a (dis-)advantage in past times gave rise to a persistent spatial equilibrium (e.g., Redding et al. (2010); Bleakley and Lin (2012); Michaels and Rauch (forthcoming); Jedwab et al. (forthcoming); Jedwab and Moradi (2016); Nunn and Puga (2012)).

Finally, our paper contributes to the literature associated with the identification of factors that drive China’s urban growth (e.g., Chan et al. (2008), Jia (2014) or Faber (2014)). Our main contribution to this branch of literature is to point out the crucial importance of fundamental factors that influenced the initial distribution of urbanization for the understanding of today’s spatial structure of urbanization.

### 3 Background

#### *Malaria in South China*

Before the start of the first eradication campaigns in the 1950s, *Plasmodium falciparum* malaria was, in contrast to North China, endemic throughout all of South China (Tang et al., 1992). Two main factors explain this regional discrepancy in the malaria environment. First, climate in subtropical (between 25 and 33 degrees North) and tropical (below 22 degrees North) South China is very conducive to the development of the *falciparum* parasite, as well as potential vectors, i.e., the mosquitoes. In the northern and western parts of China, the climate is much less suitable, if not completely unsuitable, for the transmission of *Plasmodium falciparum* malaria. Second, above 33 degrees North, typically only the *Anopheles sinensis* is found. This is a relatively inefficient vector and, crucially, does not transmit *Plasmodium falciparum* malaria. As a consequence, only the less harmful, typically non-lethal, malaria *vivax* was seasonally prevalent in North China. South of the 33rd degree North, however, there are two additional vectors: The *Anopheles anthropophagus* and the *Anopheles minimus* (Manguin et al. (2008, p.261ff.), Chow (1991)). Both are capable of transmitting the *falciparum* parasite and are more efficient than the *Anopheles*

sinensis (Manguin et al. (2008, p.261ff.), Chow (1991)).

The combined presence of the climatically suitable environment, the *Plasmodium falciparum* parasite, as well as the vectors capable of transmitting the parasite resulted in high mortality rates south of the Yangtze River, including the river valley itself. Every major city along the Yangtze was plagued by the presence of *Plasmodium falciparum* malaria in the early 20th century (Shu, 1932; Simmons et al., 1944). The large-scale, government-run malaria eradication campaigns that started in the 1950s and were intensified during the 1970s dramatically reduced the malaria-related death toll in China. The fatalities dropped from 300,000 in 1950 to 100 in 1990 (Carter and Mendis, 2002). *Plasmodium falciparum* malaria has now been eradicated in all provinces except in Yunnan and Hainan (Tang, 2000). The endemicity rate is below 5 percent in the few regions of these two provinces where *falciparum* malaria still occurs (Gething et al., 2011b).<sup>7</sup>

#### *Number of Administrative Capitals as Proxy for Economic Activity in the Imperial Era*

Local political and economic power of the Chinese state has been concentrated in the administrative capitals since the start of the unified empire. The capitals are hierarchically classified into county, prefecture and province capitals according to their ascending political and economic importance.<sup>8</sup> These capitals oversee increasingly larger regions in analogy, for example, to the U.S. administrative system of municipalities, county seats and state capitals. In the historical era, the location of the capitals typically coincided with the presence of a high population density or significant markets. This implies that the spatial distribution of administrative capitals was closely intertwined with, and representative of, the distribution of (urban) population and economic activity in the imperial era (Chang, 1961, 1963).<sup>9</sup> A noteworthy feature of the administrative regions is that they were often split into multiple units, each with its own capital, when their population densities increased significantly. A high density of administrative capitals within a given area is therefore indicative of a high population density (Skinner, 1977b, p. 231).

---

<sup>7</sup>The endemicity rate is defined as the annually averaged infection prevalence in 2–10 year olds.

<sup>8</sup>The continuity of the hierarchical system of territorial control is a remarkable feature of the Chinese state. It has remained—in its core—stable since the early days of the unified empire (Ma, 2005).

<sup>9</sup>In addition to administrative capitals, smaller towns and market places also existed during the imperial era. In comparison to the administrative capitals, however, they were only of secondary importance. In the year 1893, for example, the average administrative capital was home to 12,700 residents while the smaller urban places had a mean population of 410 (own calculations based on Table 2A in Skinner (1977b, p. 224)) The administrative capitals, on the other hand, covered the largest urban areas.



### *Malaria and Han Chinese Settlement Patterns*

The homeland of the (Han) Chinese—the ethnicity for which we can track the spatial evolution of urbanization—is the Yellow River’s central plain. It is where the first centralized Chinese state emerged under the Qin dynasty in 221 BCE. From there, the South of China was colonized in three major migration waves—triggered by natural disasters, political turmoil and warfare—over the next millennium (Holcombe (2001, p.21), Buchanan (1970, p.6ff.), Marks (2011, p.103ff.)).<sup>10</sup> This moved the center of (population) gravity from North to South China. Illustrating this shift is the fact that the share of China’s total population living in the South was 20 percent in the year 2 CE , while it had reached 71 percent in 1391. Figure 1 illustrates the southward expansion. It depicts the spatial distribution of the administrative capitals before the start of the large migration waves, after the end of the third wave of migration, and in the modern era.

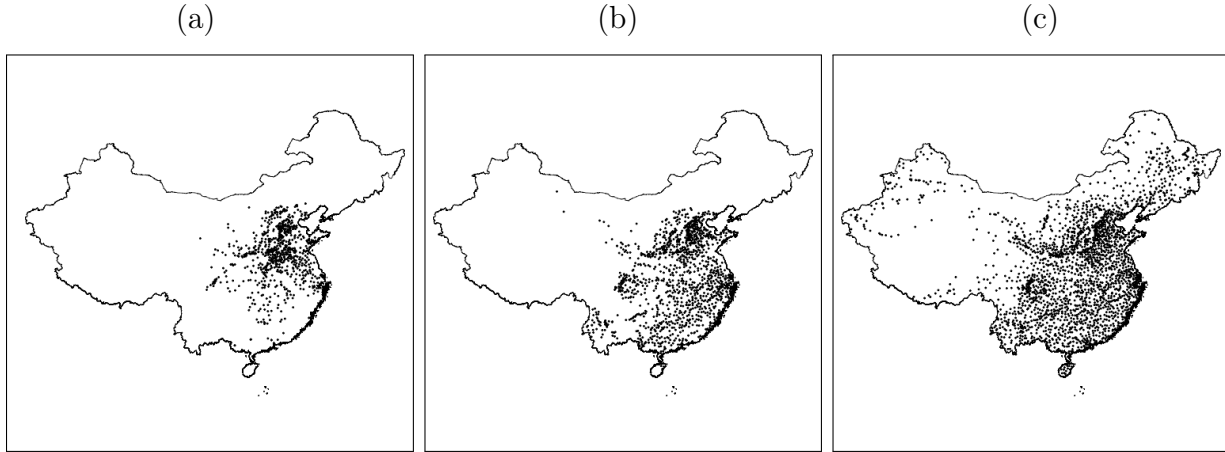


Figure 1: Spatial distribution of administrative capitals before the big southward migration waves (200 BCE, panel (a)), after the three major migration waves (1400 CE, panel (b)) and in the modern era (1990 CE, panel (c)).

Historical records indicate that settlement decisions of the Chinese colonizers were strongly influenced by the local malaria environment. Settlement in malarious areas placed a great burden on the health of the non-immune Chinese. As they were pushed southward from their non-endemic homelands, *Plasmodium falciparum* malaria caused many fatalities among the migrants (Marks, 1998, p.73ff.). The Chinese understood early on what type of geographical and climatic environments were particularly conducive to *Plasmodium falciparum* malaria transmission and should therefore be avoided (Kiple, ed (1993, p.355ff.), Lewis (2011, p.13), Zhu (2011)). We consequently expect that the settlement decisions of

---

<sup>10</sup>A more detailed discussion of the Han colonization of the South is presented in Appendix A.1.

the Han Chinese as well as their local population sizes were influenced by the local malaria environment (Wiens (1967, p.133, p.180ff.), McNeill (1976, p.79ff.), Marks (1998, p.73ff.), Marks (2011, p.129)).<sup>11</sup>

## 4 Empirical Strategy

In this section, we first describe the central properties of our measure of the climatic *Plasmodium falciparum* malaria transmission suitability. Subsequently, we discuss the regression methodology employed to analyze the effect of this measure on the spatial distribution of economic and urban centers.

### 4.1 Climate-Based Malaria Transmission Suitability Model

There are three (main) necessary conditions that have to be fulfilled in order for *Plasmodium falciparum* malaria to occur in humans: (i) The existence of a sufficiently high density of humans, (ii) the presence of both a vector, i.e., a mosquito, capable of transmitting the malaria parasite and the *Plasmodium falciparum* parasite itself, and (iii) climatic conditions that enable the development and survival of the parasite and the vector. In this study, we are interested in analyzing malaria-induced effects on outcomes related to the spatial distribution of human population. It is evident that relying on a prevalence-based measure would bias our estimates due to the inherent reverse causality between human population density and malaria prevalence. For instance, in a region where humans are absent, prevalence is zero. A second endogeneity-related concern results from the fact that the level of urbanization itself can influence malaria prevalence. For example, malaria control might be more readily implemented in cities. Furthermore, the presence of vectors and parasite can be endogenous to the spatial distribution of human population. Eradication campaigns, for example, have successfully eliminated malaria transmitting vectors. To avoid these endogeneity issues, our empirical analysis builds on a climate-based measure of *Plasmodium falciparum* malaria transmission suitability. In order to allow for the identification of the malaria-related effects, this measure has to fulfil three conditions: (a)

---

<sup>11</sup>The historical literature strongly suggests that the settlement decisions of the Han Chinese were influenced by the malaria environment. However, due to lack of data we cannot investigate whether the Han Chinese established truly new settlements or whether they co-located in existing villages of indigenous ethnicities. Relatedly, we cannot analyze whether the pre-colonial spatial population distribution was influenced by the malaria environment. What is known, however, is that population density in South China was very low prior to colonization (Wiens, 1967, p.171)

It must be exogenous with respect to the geographical distribution of human population; (b) it has to be specifically modelled to capture the suitability for *Plasmodium falciparum* malaria, i.e., the most harmful type of malaria, and (c) the model has to feature variation across space.<sup>12</sup>

A measure that fulfils the three criteria mentioned above can be derived from the ecological model of *Plasmodium falciparum* malaria transmission intensity developed in Mordecai et al. (2013). This model describes the mosquito as well as the parasite development as non-linear functions of air temperature. Due to their thermal physiology, air temperature influences both the mosquito life cycle and the malaria parasite development rate. Consequently, the spatial extent as well as the intensity of malaria transmission is crucially influenced by temperature. Our modified version of the Mordecai et al. (2013) model consists of three multiplicative components, each representing a non-linear function of air temperature, that describe the mosquito density, the mosquito mortality rate, and the extrinsic incubation period of the *Plasmodium falciparum* parasite in the mosquito, respectively.<sup>13,14</sup> Of critical importance for our study is the fact that the parametrization and the functional forms of these individual components are based solely on experimental data from laboratory studies that analyze the thermal physiology of the parasite and the vector. The results of these studies are—unlike values derived from field studies—independent of local differences in human population densities. The functional forms and parameter values resulting from the laboratory studies are gathered in Mordecai et al. (2013).<sup>15</sup> We adopt them directly.

---

<sup>12</sup>The malaria intensity measure that is predominantly used in the field of economics is the global malaria stability index developed in Kiszewski et al. (2004) (see, for example, Nunn and Wantchekon (2011), Henderson et al. (2012), or Alsan (2015)). This index does not fulfil these criteria and therefore is not suited to the purpose of our study. The measure is a function of the observed human biting rate—that is, the proportion of mosquitoes that bite humans rather than animals—of the observed locally dominant mosquito vector. Data on both, the dominant vector and biting rate, were collected in field studies and are endogenous with respect to human population density (e.g., Smith and Ruktanonchai (2010), Ngwa (2006) or Killeen et al. (2001)). Similarly, the malaria index used in Gallup and Sachs (2001) is directly derived from observed malaria prevalence and observed human population distribution patterns.

<sup>13</sup>The functions are either quadratic functions or left-skewed unimodal functions of air temperature (see Appendix A.2 for details).

<sup>14</sup>Mordecai et al. (2013) additionally incorporate the human biting rate of the *Anopheles*, the proportion of the bites of infective mosquitoes that infect susceptible humans, and the proportion of bites of susceptible mosquitoes on infectious humans that infect mosquitoes as temperature-dependent components in their measure. Because these three components are potentially affected by the co-evolution of humans and vectors we hold them constant in our measure of malaria suitability. That is, we do not describe these components as functions of air temperature. However, our empirical results remain stable if we substitute the original Mordecai et al. (2013) model for our malaria suitability measure (Table C.10).

<sup>15</sup>The functions together with their parameter values are depicted in Table 2 in Mordecai et al. (2013).

Our malaria suitability measure can be interpreted as representing a natural locational fundamental. This fundamental could have a negative effect on the human population density through two directly health-related channels. First, (non-immune) humans that settle in a climatically suitable region could fall ill to malaria and die or relocate. Second, given the knowledge of areas with a high malaria transmission potential, migration and settlement by non-immune people into these regions is deterred. In this case, we would observe an impact of the climatic malaria suitability on the settlement pattern even in the absence of malaria prevalence. People do not themselves have to experience a health burden in order to avoid climatically suitable areas for malaria transmission. In addition to these effects related to contemporaneous health, the local malaria potential can also influence local population densities and economic development through its past effect. As outlined in the Section 1, permanently adverse health conditions can influence long-term determinants of growth, such as local institutions or social and cultural traits (e.g., Acemoglu et al. (2001)). Furthermore, when increasing returns exist, initial malaria-influenced decisions on where to establish urban centers can have persistent effects on the spatial pattern of urbanization and economic development due to the benefits associated with concentration of population and economic activity in already existing centers.

Using high-resolution temperature data, we compute our malaria transmission intensity measure for each  $0.5 \times 0.5$  degree grid cell that lies within the territory of today's China. The high-resolution spatial data, reported at monthly intervals, is available starting with the year 1901. To obtain a time-invariant measure of the local climatic malaria potential, we take the average of the monthly malaria suitability index over the period 1901–1950. This represents the time span between the start of the availability of the climate data (1901) and the beginning of the malaria falciparum eradication campaigns. By using this time-invariant measure, we assume that the relative position of a given cell in the malaria suitability distribution remains constant over the time period of our empirical analysis. In Appendix A.2, we provide evidence that this assumption is plausibly fulfilled. The Appendix also contains a more detailed description of our malaria suitability measure.

## 4.2 Methodology

We conduct our empirical analysis at two levels of aggregation: For the main analysis, we construct a grid-cell level dataset. In supporting regressions, we additionally make use of county-level data. Below, we illustrate our empirical approach at the grid-cell level. The

methodology translates directly to the county level.

To analyze the effect of malaria suitability on the spatial distribution of population and economic activity, we employ the following linear cross-section regression model:

$$y_c = \theta MSM_c + \beta' \mathbf{X}_c + \varepsilon_c. \quad (1)$$

The various dependent variables of grid cell  $c$  are represented by  $y_c$ . Our measure of the grid-cell-level climatic malaria suitability is given by  $MSM_c$ . The vector  $\mathbf{X}_c$  contains exogenous geographical and climatic variables that are commonly associated with urbanization and economic development. Specifically, this set of controls contains the shortest distance from the center of the grid cell to the coastline, the mean elevation of the cell, as well as the Strahler index representing the cell-specific accessibility by waterways. Due to the particular importance of the Pearl and Yangtze rivers with respect to agricultural development and transport, we include two dummy variables indicating whether the respective river runs through a given grid cell. As an additional control for the importance of waterway networks, we include river-drainage fixed effects. Further, the agricultural potential of each cell is captured by an agricultural suitability index. We also control for the land mass of the grid cell. In all our regressions, we include fixed effects for the physiographic macroregions (Skinner, 1977b). The areas are defined according to geographical features that structured the socio-economic landscape of China prior to the advent of modern transportation.<sup>16</sup>

Additionally, we control for temperature and precipitation, as well as their respective squared values and first-order interactions. The climate controls account for the possibility that socio-economic variables are directly influenced by climatic conditions (e.g., Gallup et al. (1999), Nordhaus (2006), Barrios et al. (2010) or Dell et al. (2012)). All of the climatic variables are averaged over the period 1901–1950. As further climate-related controls, we include the absolute longitude and latitude as well as climate-zone fixed effects in our regression setup. Note that as we include air temperature (and its squared term) directly as control variables in the regressions, our identifying variation is the residual variation between our malaria-falciparum-specific function of air temperature and the direct temperature effects.

With respect to statistical inference, we cluster the standard errors at 2×2 degree ‘super

---

<sup>16</sup>Our results remain qualitatively equivalent if we replace the macroregions with the first-level administrative areas, i.e., provinces. In contrast to the former, however, the latter are not necessarily exogenous with respect to our malaria transmission measure due to the fact that the location of administrative borders can be dependent on human population densities.

grids’ (cf., Motamed et al. (2014)). In Appendix C.2, we show that we obtain very similar standard errors when we employ the approach of Conley (1999) to account for spatial correlation or cluster the standard errors at the province level.

The identifying assumption in Eq.(1) is that, conditional on covariates  $\mathbf{X}_c$ , the MSM affects our outcome variables only through changes in malaria suitability and is unrelated to any other factors in the error term. Potential threats regarding this assumption are discussed in Section 6.

## 5 Data and Descriptive Analysis

This section provides an overview of the data used in the empirical analysis along with a descriptive analysis thereof.

### 5.1 Data

For our analysis, we construct two datasets. Our main dataset is the grid-cell level dataset in which we divide the territory of current-day South China into 868 equally sized cells of  $0.5 \times 0.5$  degrees.<sup>17</sup> The dividing line between North and South China is drawn along the Huai River–Qin Mountains line. Building on the grid-cell level data, we construct a secondary dataset at the county level.

#### *Grid-Cell Level Dataset*

For each of the 868 grid cells, we compute the time-invariant malaria suitability measure, as well as the mean temperature, and the mean precipitation by taking their respective averages over the period 1901–1950. The monthly climate data are provided by the Climatic Research Unit at University of East Anglia. Further, we determine the agricultural suitability (Ramankutty et al., 2002), the mean elevation (NOAA’s National Geophysical Data Center) and the distance to the nearest coastline for each cell. To capture the accessibility by waterways, we use the Strahler order of streams. This is an index that ranges from one to six, representing the size of the largest river (Vörösmarty et al., 2000). To account for their crucial importance, we define two separate dummy variables that indicate

---

<sup>17</sup>In Appendices C.1 and C.3, we will also run regressions on a sample of grid cells that lie in North China.

whether a grid cell is intersected by the Yangtze or the Pearl river.<sup>18</sup> This information is extracted from [www.naturalearthdata.com](http://www.naturalearthdata.com). The information with respect to the river drainage basins, macroregions and climate zones stem from the WaterBase Project<sup>19</sup>, Skinner (1977b) and Rubel and Kottek (2010), respectively. We also control for the size of the land mass encompassed by each grid cell which changes with its geographical position. To analyze the spatial distribution of administrative capitals over time, we use information contained in the CHGIS (2010) database.<sup>20</sup> This data source reports the geographical coordinates of the administrative centers over the time period 221 BCE to 1911 CE at an annual interval. For the analysis at the grid-cell level, we define the number of seats located within each grid cell as the measure for urban population density during the imperial era (see Section 3). In addition to the location of the administrative seats, Skinner et al. (2011) provide information on their population numbers as well as qualitative characteristics for the year 1893, i.e., the late imperial period. Information with respect to population sizes of administrative capitals in 1893 is reported in bins. Specifically, the capitals are categorized into one of the following 12 bins: [0, 500), [500, 1,000), [1,000, 2,000), [2,000, 4,000), [4,000, 8,000), [8,000, 16,000), [16,000, 32,000), [32,000, 64,000), [64,000, 128,000), [128,000, 256,000), [256,000, 512,000) and [512,000,  $\infty$ ). We assign each seat a population corresponding to the mean of the bin into which it falls. For example, all seats associated with the bin [16,000, 32,000) are assigned a population of 24,000. The exception is the largest category of cities to which we assign a population of 512,000.<sup>21</sup>

In addition to the population variable, we define an ordered variable that represents the administrative level of the highest ranked city located within a grid cell. This hierarchy-level variable takes integer values ranging from zero to three. Cells with no administrative seats are assigned the value zero. Grid cells containing county, prefecture and province seats, respectively, are assigned the value one, two and three.

To determine the grid-cell-level urban population in the modern era, we use two data

---

<sup>18</sup>The Yellow river also constitutes a crucially important waterway. However, it lies entirely in North China.

<sup>19</sup>[www.waterbase.org](http://www.waterbase.org)

<sup>20</sup>More specifically, we include the county, prefecture and province seats that are part of the standard administrative system. The standard administrative system encompasses all seats which are under direct control of the empire. This excludes, for example, seats that were occupied (governed) by local chieftainships (Skinner, 1977a, p.304). On average, 95 percent of all administrative capitals are part of the standard centralized system.

<sup>21</sup>The limited information with respect to the city size imposes two main restrictions: First, it reduces the variation in the population variable exploitable in the regression analysis. Second, the population number of large cities is capped at 512,000, which truncates the city-size distribution from above. These limitations should be kept in mind when interpreting our results.

sources, each representing a different point in time. Both databases are constructed on the basis of official Chinese population censuses. The first data source is Skinner et al. (2013). This database covers the universe of urban places in China. In total, there are 7,225 places in South China that are classified as ‘urban’. Within these, the number of urban residents ranges from 57 to 7,353,000; the median is 2,486. This indicates that many of the urban places are small. We want to stress that our regression results remain stable if we apply a minimum cut-off level to the population data.<sup>22</sup> To compute the number of urban dwellers residing within each grid cell, we simply sum over all urban places that are located within a given cell.<sup>23</sup> The second data source used to determine the grid-cell urban population is Brinkhoff (2015). This database is less extensive than the Skinner et al. (2013) data. It only reports the number of residents of cities that have more than 75,000 inhabitants in the year 2010.<sup>24,25</sup>

Finally, we compute the grid-cell number of workers employed in the manufacturing sector. The raw data is reported at the county level. Using the Skinner et al. (2013) data, we can assign the county-level data to the individual grid cells. This assignment procedure is described in Appendix B.

An important note relates to the definition of our dependent variables. We do not use the raw numbers in the regressions, but instead define the variables as percentages of the respective mean. For example, the number of administrative capitals located within a grid cell is divided by the average grid-cell number of capitals in a given year. Likewise, the grid cell’s urban population is expressed relative to the overall cell-level mean in a particular year. This definition accounts for the average degree of urbanization which changes over the years and, consequently, allows us to compare the coefficient estimates across time periods (e.g., Henderson (2005)). The resulting point estimates can be interpreted as the change in the dependent variable evaluated at the sample mean. For example, when analyzing the effect of our malaria suitability measure on the urban population, the coefficient represents

---

<sup>22</sup>We experimented with various population cut-offs, including 5,000, 10,000 and 75,000. In all setups, our results remained qualitatively stable.

<sup>23</sup>A possible concern with this approach is that we neglect potentially important aspects of urban sprawling due to the fact that the Skinner et al. (2013) population data represents points rather than areas. However, as presented in Appendix C.2, we reproduce our results using the population data of Klein Goldewijk et al. (2010) which contains information with respect to the extent of urban areas.

<sup>24</sup>In addition to the year 2010, the database also contains population numbers for the year 1990 and 2000. We are able to reproduce our results obtained for the year 2010 and presented in Section 6 for these two alternative years.

<sup>25</sup>The Brinkhoff (2015) data are not geocoded. To each city, the coordinates of its center were assigned manually.



the reduction in the number of grid-cell-level urban inhabitants—expressed as percentage of the average urban population in that year—that is associated with a one-standard deviation increase in malaria transmission suitability. Note that this is equivalent to evaluating the change in a grid cell’s share of total urban population at the average share that results from a one-standard deviation increase in the malaria transmission intensity.<sup>26</sup>

### *County-Level Dataset*

To shed further light on modern-day regional disparities induced by differences in the malaria environment, we employ the county-level data from Skinner and Henderson (2012). For the year 1990, this data source provides detailed information regarding income, employment structure, and demographic characteristics of the population. On the basis of the county polygons provided in Skinner and Henderson (2012), we construct the county-level climate and geographical variables by taking the mean of the grid cells over which a given county spans.

## **5.2 Descriptive Analysis**

The purpose of this section is to (a) present summary statistics of our key variables, and (b) provide evidence that our climatic malaria transmission intensity measure successfully captures the (potential) local malaria burden.

### *Summary Statistics*

Table 1 depicts the summary statistics of the key variables of our grid-cell level dataset. The explanatory variables that are constructed on the basis of climate data represent averages over the period 1901–1950. The values of the outcome variables are gathered for different years as indicated in the table. The summary statistics used in the supporting county-level regressions are presented in Appendix A, Table B.1.

The variables labeled ‘Population’ represent the number of inhabitants that resided in urban areas within a given grid cell and year as a percentage of the average urban population

---

<sup>26</sup>That is, the number of urban residents expressed as a percentage of the average is the same as the share of total urban residents relative to the average share. More generally, note that the number of urban residents, the number divided by the average, as well as the number of the grid cell urban inhabitants divided by the total urban population are collinear. The only difference is the scaling of the variables. Consequently, the results shown below remain qualitatively equivalent irrespective of which of the three definitions is employed. The only difference lies in the interpretation of the coefficient estimates which changes according to the scaling.

Table 1: Descriptive Statistics Key Variables

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
Population <sub>1893</sub>	1	2.900	0	25.334	868
Population <sub>1990</sub>	1	3.265	0	69.924	868
Population <sub>2010</sub>	1	4.111	0	70.317	868
Manufacturing Employment <sub>1990</sub>	1	2.621	0	50.724	868
Malaria Suitability (SD)	0	1	-2.526	3.489	868

in the respective year. The standard deviations of the three population variables are very similar. The relatively low maximum value in the year 1893 is (partly) the result of the size of big cities being capped. The grid-cell-level industrial employment for the year 1990 as a percentage of average industrial employment in that year is captured by the variable ‘Manufacturing Employment’.

#### *Malaria Suitability and Malaria Prevalence*

To investigate the relationships between our malaria suitability measure, malaria prevalence and human population density, we geocode information on the location of malaria falciparum cases in the pre-eradication era reported in Tang et al. (1992).<sup>27</sup> This allows us to determine in which grid-cells malaria was prevalent, i.e., for which areas cases of malaria falciparum were reported.<sup>28</sup>

Column (1) of Table 2 depicts the correlation between our measure of the climatic Plasmodium falciparum malaria transmission potential and the probability that this type of malaria was prevalent in a given grid cell. As expected, this correlation is positive, indicating that our measure, in fact, represents a locational fundamental that adversely affects health. The higher the suitability, the higher is the probability that malaria is prevalent in a region. We further observe a positive association between malaria prevalence and urban population density (column (2)–(3)).<sup>29</sup> This positive correlation documents the existing

<sup>27</sup>Figure B.2 in the Appendix A depicts the spatial distribution of malaria occurrence in the pre-eradication era.

<sup>28</sup>Note that this variable does not capture the intensity of malaria transmission. Rather, it constitutes a dummy variable that indicates whether Plasmodium falciparum cases were observed or not. This restriction is imposed by the nature of the historical prevalence data which only mentions the location of malaria falciparum cases but does not contain information on to the intensity of malaria transmission.

<sup>29</sup>We have no information on population concentrations around the year 1950. However, the fact that the (urban) population concentration for the year 1893 as well as that for 1990 are correlated with the pre-eradication prevalence together with the fact that the spatial distribution of urbanization is stable over time (see Appendix B.1) implies that the observed results also pertain to the year 1950.

Table 2: Correlation: Malaria Suitability, Malaria Prevalence and Population Density

Dependent Variable:	Pre-Eradication Malaria Prevalence			Urban Population 1893	
	(1)	(2)	(3)	(4)	(5)
MSM (SD)	0.132** (0.059)			-0.821*** (0.208)	
Population <sub>1893</sub>		0.013* (0.007)			
Population <sub>1990</sub>			0.011*** (0.003)		
Pre-Eradication Malaria Prevalence					0.462 (0.318)
Obs.	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Standard errors are clustered at  $2 \times 2$  degree grid cells reported in parentheses. Estimates represent correlation coefficients conditional on absolute latitude, temperature, temperature squared and area of the grid cell.

reverse causality between malaria prevalence and human population density. Evidence for the severity of the endogeneity issue associated with the use of prevalence-based measures is provided in columns (4)–(5). The relationship between the share of total urban population residing within a grid cell in the year 1893 and our exogenous climatic malaria suitability measure is strongly negative and statistically highly significant.<sup>30</sup> On the other hand, there is a positive relationship between population numbers and malaria prevalence in the year 1950 (which is borderlining the 90 percent confidence level). This indicates that regressions that are based on prevalence-derived malaria intensity measures will produce coefficient estimates that are biased towards zero. By relying on our exogenous measure of the climatic potential of transmission we avoid this issue.

## 6 Results

In this section, we first document that the climatic malaria transmission suitability represented a locational fundamental that strongly and continuously influenced the spatial distribution of urban centers of the Han Chinese from the start of their southward expansion up to the end of the era of imperial rule. This first step thus covers a time period of more than 2000 years during all of which malaria falciparum represented a major health risk. In a second step, we direct our attention to the current-day distribution of urbanization and economic activity. We analyze whether malaria-related effects persist after

<sup>30</sup>We obtain very similar results if we use population data for the year 1990 or 2010.

eradication.

## 6.1 Malaria Suitability and Urbanization in the Imperial Era

To investigate the effect of malaria environment on the spatial distribution of urbanization during the imperial era, we run the cross-section regression model of Eq.(1), with the number of administrative seats (expressed as a percentage of the mean) as the dependent variable, at an interval of 100 years over the period 200 BCE to 1900 CE. As outlined in Section 3, the number of administrative capitals located within a grid cell reflects the urban population density during the imperial era.

The results of the 22 separate regressions are depicted in Figure 2. The circles represent the point estimates, the bars indicate the 90 percent confidence intervals, where the standard errors are clustered at  $2 \times 2$  degree grids. In addition to the coefficient estimates, Figure 2 depicts the number of administrative capitals located in South China (dashed line). The successive colonization of the South is reflected in the positive trend in the number of cities between 200 BCE and 700 CE. After that, the number of administrative seats remains relatively stable until 1900, the last year observed in the imperial era.

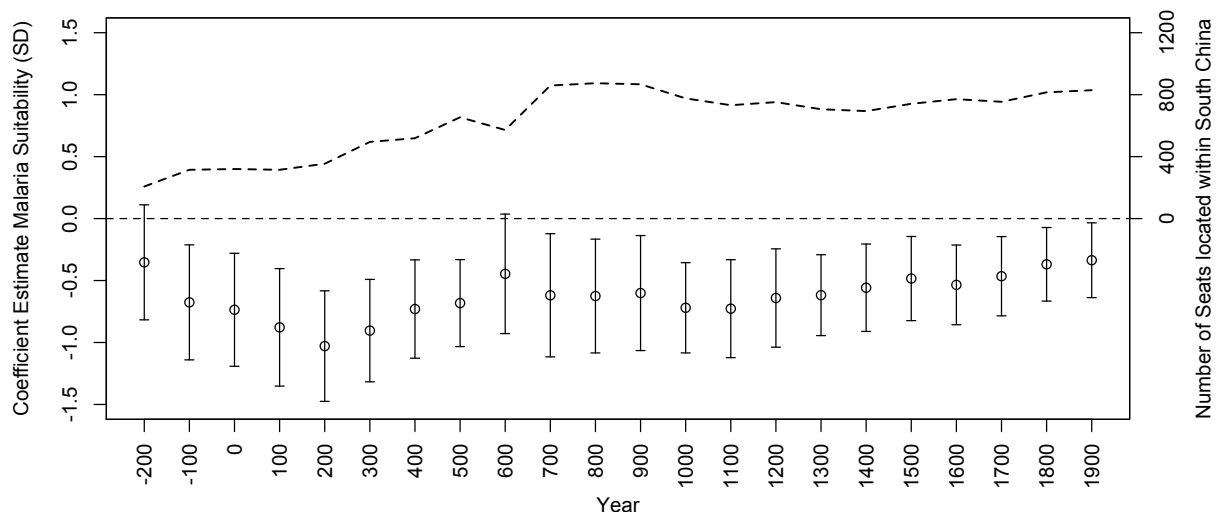


Figure 2: Graphic representation of the coefficient estimates of the MSM effect on the number of administrative capitals located within a grid cell (evaluated at the sample mean) at different points in time. The table of results is shown in Appendix C.2, Table C.1 Panel A.

The coefficient estimates depicted in Figure 2 document that the local climatic malaria transmission potential constituted a negative locational fundamental that strongly influ-

enced the spatial distribution of the administrative seats during the imperial era.<sup>31</sup> The point estimates of our malaria suitability measure are negative in all years and significantly so shortly after the start of the unified empire and its expansion into South China (100 BCE). The variation in the size of the point estimates is likely a reflection of changes in the territorial extent of the empire at different points in time. For example, an explanation for the somewhat higher (closer to zero) and noisier estimate for the year 600 is the fact that this observation falls within the period of the Sui dynasty during which large parts of southwestern China were not under imperial control. Generally, however, the coefficient estimates are relatively stable, ranging from  $-0.26$  to  $-0.93$ . The average coefficient size of  $-0.6$  implies that the number of seats, evaluated at the sample mean, is reduced by 60 percent when moving from a given area to a region characterized by a one-standard deviation higher climatic malaria transmission suitability.

In Appendix C.1 Table C.2, we show that the results of Figure 2 remain stable if we include only a very limited set of climatic control variables. This also holds true when we successively add more controls (Table C.2). Additionally, we re-run the 22 regressions on a sample that encompasses both the southern and the northern parts of China Proper (Table C.1, panel B). While the general pattern of results remains stable, the point coefficients are less precisely estimated and, on average, reduced in size by more than 50 percent. This can be explained by the fact that our climatic malaria suitability measure did not represent a health burden in North China due to the absence of a vector capable of transmitting the *Plasmodium falciparum* parasite (see Section 3). This notion is confirmed by the null results obtained from regressions in which we restrict the sample to North China (Table C.1 panel C). There is neither a clear pattern in the sign of the malaria suitability coefficients nor are they statistically significant. Variation in the malaria suitability measure stemming from North China consequently introduces noise in our estimates without being informative about the adversity of the local disease burden.

In the next step, we show that the malaria transmission potential affected city characteristics other than location alone. To this end, we employ data collected by Skinner et al. (2011). This data source contains qualitative information on all administrative capitals in the year 1893. In the first column of Table 3, we confirm that the general results of Figure

---

<sup>31</sup>Due to the lack of data on the spatial distribution of population in South China prior to Han colonization, we cannot exclude the possibility that the Chinese settlement pattern reflects pre-existing, malaria-influenced settlement patterns of indigenous ethnicities. We therefore do not claim that the settlement behavior of the Han Chinese created an entirely new spatial distribution of human population. The data simply tell us that since the early start of Han colonization of the South, the location of their administrative capitals has been influenced by the local malaria environment.

2 also hold for the year 1893. Fewer administrative capitals are located in grid cells that are characterized by a high degree of malaria transmission suitability. Along with a grid cell’s number of capitals its share of total urban population is reduced (column (2)). Moving from a given area to a region characterized by a one-standard deviation higher degree of malaria transmission suitability reduces the share of total urban population living in a grid cell, evaluated at the sample mean, by 74 percent. The findings presented in columns (1)–(2) are also reflected in the hierarchy level. Column (3) shows that the maximum level of administrative status attained by the capitals located within a grid cell decreases as the climatic malaria transmission potential increases.<sup>32</sup> This result confirms the existence of a close link between the size of urban population and the ranks of the cities (cf. Abramson (2006)).

Table 3: Malaria Suitability and Qualitative Features of Administrative Seats in 1893

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1893
	(1)	(2)	(3)	(4)
MSM (SD)	-0.428*** (0.160)	-0.744** (0.104)	-0.457** (0.104)	
Number of Seats <sub>1893</sub>				1.053*** (0.176)
Climate Controls	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes
R-squared	0.190	0.142	0.141	0.250
Obs.	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

In the regressions presented in Figure 2, we have—based on the historical literature—equated the number of administrative capitals located within a grid cell with its urban population density. Using the Skinner et al. (2011) data we can show that this assumption plausibly holds. Column (4) of Table 3 documents that the relationship between the number of seats and the urban population is close to unity in the year 1893. This result strengthens our previous conclusion that the local potential malaria transmission intensity

<sup>32</sup>This result remains unaltered if we explicitly take into account the ordered—and possibly non-linearly so—nature of the hierarchy variable in an ordered logit regression setup (Appendix C.2, Table C.3, column (1))

strongly and continuously influenced the spatial distribution of urbanization during the whole period of imperial rule, i.e., a time period in which malaria falciparum constituted a major health risk. In the remainder of the paper, we turn our attention to the present era, i.e., the post-eradication era.

## 6.2 Malaria Suitability and Urbanization in the Modern Era

In Table 4 we investigate whether regions with a high climatic malaria transmission potential are still characterized by a lower level of urbanization and economic activity in the modern era. In column (1) we use detailed data of Skinner et al. (2013) that covers the universe of urban places in the year 1990 to document that the local degree of malaria transmission suitability continues to exert a strong negative effect on urban population. Evaluated at the sample mean, a one-standard deviation increase in the MSM reduces the share of the total urban population located within a grid cell by 72 percent. This result documents that the malaria environment continued to influence the spatial pattern of urbanization until 1990. A number of major reforms were implemented in China between 1990 and 2010, including large waves of privatization and the relaxation of labor mobility restrictions (e.g., Yang and Yu (2011); Bosker et al. (2012)). To investigate whether the malaria environment continues to be an important determinant of the spatial distribution of urbanization after these reforms, we employ population data for the year 2010, drawn from Brinkhoff (2015). As depicted in column (2) of Table 4, the point estimate of  $-0.808$  is very similar to the coefficient obtained using the 1990 data. Furthermore, the comparison between the malaria suitability effect in the modern and the late imperial era (Table 4 column (2) versus Table 3 column (2)) shows that they are very similar and statistically indistinguishable.<sup>33</sup> This suggests that the reduced-form effect of the malaria environment has remained unchanged between 1893 and 1990 (2010). We do not observe regional convergence between high and low malaria transmission potential areas over the period 1893–1990 even though the climatic malaria environment has ceased to represent a locational disadvantage in the latter half of this period.<sup>34</sup> When interpreting these results it is important to recall that our estimates represent reduced-form effects. While the results of columns (1)–(2) document that the local climatic *Plasmodium falciparum* transmission

---

<sup>33</sup>The p-value for the test of coefficient equality of the MSM effect in 1893 and 1990 is 0.855 and 0.713 for the test of equality of the effect in 1893 and 2010.

<sup>34</sup>The degree of persistence in the spatial distribution of urbanization between 1893 and 1990 is discussed in more detail in Appendix B.1.

Table 4: Malaria Suitability, Urbanization and Industrial Activity

Dependent Variable	Urban Population 1990	Urban Population 2010	Manufacturing Employment 1990
	(1)	(2)	(3)
MSM (SD)	-0.719*** (0.229)	-0.808*** (0.299)	-0.631*** (0.204)
Climate Controls	yes	yes	yes
Geography Controls	yes	yes	yes
R-squared	0.132	0.139	0.287
Obs.	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

potential has a substantial negative effect on current-day urbanization, we cannot derive any insights regarding the underlying mechanisms.

As outlined in Sections 1 and 4.1, there are a variety of channels through which the malaria environment can influence current-day outcomes. One way is via its effects on long-term determinants of urbanization and economic development. For example, the adverse health environment can influence the evolution of local institutions or cultural traits. Furthermore, it can influence the initial location of urban centers, as evidenced by our earlier results. The existence of local increasing returns and coordination problems imply that geographic characteristics that were important during the initial phases of urbanization can permanently influence the spatial pattern of urbanization even when these characteristics have since long become obsolete (e.g., Davis and Weinstein (2002); Bleakley and Lin (2012); Michaels and Rauch (forthcoming)). The persistent effect of the malaria environment suggests the existence of considerable local increasing returns to scale in South China. One important, China-specific, returns-to-scale-related aspect that potentially reinforced existing urbanization pattern is the post-revolutionary industrialization strategy of the Chinese state that disproportionately directed investments towards large cities (Chan et al., 2008; Chan, 2010). Urban industrialization created employment opportunities which have, in turn, led to a dramatic increase in the urbanization rate. The close interrelationship between urbanization and industrialization is mirrored in the estimate presented in column (3) of Table 4. We find a statistically significant negative effect of the malaria environment on industrial activity as proxied by manufacturing employment.<sup>35</sup> Moving from

<sup>35</sup>We do not possess information regarding the spatial distribution of manufacturing activity for the year



a given area to a region that exhibits a one-standard deviation higher malaria potential reduces the grid cell’s share of total manufacturing employment, evaluated at the sample mean, by 63 percent. Compared to malaria suitability’s effect on urbanization (columns (1)–(2)) the size of the coefficient is very similar and statistically indistinguishable. Based on the fact that there is a large urban-rural income gap in China<sup>36</sup>, we expect that the local degree of malaria transmission suitability will not only explain variation in urbanization and industrialization across space but also in income per capita. In Section 6.3 we provide empirical evidence that this is the case. Furthermore, we provide suggestive evidence for the existence of increasing returns to scale by showing that regions with a low climatic malaria transmission potential experience disproportionate in-migration.

In addition to its effect on long-term determinants of urbanization and economic development, the malaria environment can also influence today’s outcomes via its direct effect on contemporaneous health. Specifically, eradication of the disease—implying that the climatic malaria transmission potential no longer represents a health risk—resulted in considerable health gains. For example, malaria-related fatalities dropped from 300,000 in 1950 to 100 in 1990 (Carter and Mendis, 2002). Based on theoretical and empirical results (e.g., Bleakley (2010) or Weil (2014, p.31)), we would expect that these improvements in health increase productivity and lead to income growth.<sup>37</sup> There are three potentially important explanations for why convergence between high and low malaria transmission potential is not observable in the presence of eradication-induced health gains.

First, health-related gains take time to translate into economic growth. For example, human capital of adults—the population group engaged in the production process—is primarily determined by health conditions during infancy (e.g., Bleakley (2007); Ashraf et al. (2009); Maccini and Yang (2009)). Full benefits of malaria eradication are unlikely to materialize before the labor force entirely consists of post-eradication cohorts. Similarly, a delayed effect on growth is possible when an eradication-induced increase in life expectancy raises investment in education (Kalemli-Ozcan et al., 2000; Matteo Cervellati, 2005) but this accumulation of human capital takes time to translate into economic growth due to a low initial level of productivity (Matteo Cervellati, 2005). The slow pace of these processes that relate improvements in health to economic growth could explain the persistent effect of

---

2010. We therefore cannot rerun the regressions of column (3) for the year 2010.

<sup>36</sup>For example, Chan et al. (2008) document that income per capita is 2.4 times higher in urban than in rural areas. They additionally show that average income per capita is increasing in city size.

<sup>37</sup>To empirically analyse within-region effects of malaria eradication campaigns, we need spatially disaggregated information on health and economic conditions for pre-and post eradication periods. Such data, however, is not available (Meng et al., 2015).

the malaria environment. In the short run, it is even possible that reductions in mortality depress income per capita if the drop in mortality cause population to grow (Ashraf et al., 2009; Acemoglu and Johnson, 2007; Lucas, 2013). This is the case, for example, if health-induced transition processes, such as the quality-quantity trade-off (Kalemli-Ozcan, 2002; Soares, 2005), are sluggish. In the case of China, however, population dynamics after malaria eradication are likely to be atypical due to the introduction of the one-child policy in 1978. In fact, using county-level data below, we provide evidence that mechanisms working via increased population growth are unlikely to underlie our negative reduced-form effects of the malaria environment.

The second potential explanation for the non-observance of convergence is that the effects of the malaria environment on long-term determinants of economic growth are (relatively) more important than its effect on contemporaneous health.<sup>38</sup> The latter effects cease to exist after eradication while the former continue to influence development. Eradication leads to large health-related improvements in economic conditions in previously highly malarious areas. But these regions may nevertheless fail to catch up if other malaria-environment-influenced factors that generate economic growth are more important. Increasing returns, for example, continuously reinforce the existing malaria-influenced spatial distribution of urbanization. In our county-level analysis we provide suggestive evidence for the importance of local increasing returns to scale in explaining the observed persistence.

Finally, a third possibility is that convergence in regional development is already under way but not detectable over the short period of 50 years that has passed since the beginning of the malaria eradication campaigns.

Summing up, there are a variety of potential explanations for the observed persistence in the malaria-influenced spatial distribution of urbanization in South China. This suggests that geography, in our case the climatic malaria environment, can affect today's outcomes via a multitude of complex mechanisms. In fact, it is conceivable that all of the three above-mentioned channels contribute to the explanation of our results. Given that our analysis is restricted to one single country and abstracts from differences in national institutions, our results highlight the localized nature of geography-related effects.

---

<sup>38</sup>Acemoglu et al. (2003), for example, argue that the direct effects of disease environments on contemporaneous health explain only a small fraction of existing income differences across countries and that indirect effects working via institutions are more dominant. They further note that empirical evidence that document that eradication of diseases leads to rapid economic growth is weak.

### *Robustness Checks*

The identifying assumption in Eq.(1) is that, conditional on covariates, the MSM affects our outcome variables only through variation in the degree of malaria suitability while being unrelated to any other factors in the error term. There are several potential concerns pertaining to this assumption. Below we discuss these concerns.

An immediate concern is that the MSM picks up malaria-unrelated climate effects on economic activity or settlement patterns. Based on the non-monotonicity of our index, combined with the inclusion of climate control variables in our regression setup, we do not expect this to be the case. The controls capture all direct effects of the climatic variables on our LHS variables that do not explicitly follow the same functional form as our malaria suitability model. Furthermore, in Flückiger and Ludwig (2016) we demonstrate that the MSM is unrelated to economic activity or population distribution in areas that are climatically very suitable for the transmission of malaria, but in which the *Plasmodium falciparum* parasite was not, or only weakly present. Similarly, we do not find any significant effect of our malaria transmission measure if we re-run our regressions for North China, i.e., a region in which a vector capable of transmitting *Plasmodium falciparum* malaria is (was) largely absent (Tables C.1 and C.17). These results document that our measure does not represent a negative locational fundamental in the absence of the *Plasmodium falciparum* parasite or a suitable vector. A closely related concern is that our results might be due to our malaria suitability measure being correlated with other important (and omitted) locational fundamentals that influence the spatial distribution of urbanization. While we control for the most obvious factors in the regressions presented in the main text, such as agricultural suitability or climate, we additionally show that our coefficient estimates remain stable irrespective of the inclusion of fundamentals that have been suggested by the historical literature on China (e.g., Lewis (2009)) of having been very important for population growth (Tables C.4–C.8). These variables include: Rice suitability, wheat suitability, tea suitability soil type and the presence of minerals (taken into account by including geological province dummies).

A second potential issue is that our results might be specific to our malaria suitability measure. However, as shown in Appendix C.2, Table C.9 we obtain similar estimates if we use an alternative index of the climatic malaria suitability developed by Gething et al. (2011a) or the original measure of Mordecai et al. (2013) (Table C.10). Similarly, we show that our results are reproducible using alternative population data from Klein Goldewijk et al. (2010).

Another concern is that our results are driven by the choice of the time span used to construct the average, time-invariant, malaria suitability measure and the climate variables. This is unlikely to be the case, as shown in Appendix C.2. Our results remain stable irrespective of the years included to compute the malaria measure and the climatic variables. This indicates that the relative differences in the malaria suitability among the cells, i.e., our identifying variation, have remained stable over time. In Appendix A.2, we provide more formal support for this assumption. We show that even though China experienced various periods of marked temperature variations over the last two millennia, the (relative) position of a given region in the distribution of the malaria transmission suitability is very stable over time. This is further implied by the extremely strong serial correlation observed in the temperature and MSM time series. Finally, we show that our results remain stable when we include only a very limited set of climate control variables (Table C.12).

As a final step in our analysis, we investigate to what extent the malaria-related effects on the spatial distribution of urbanization are also reflected in regional income disparities. To this end, we switch from the grid-cell level to the county level.

### 6.3 Malaria Suitability and Disparities at the County Level

A big issue related to the cross-sectional analysis on the basis of county-level data is that the local climatic potential for malaria transmission influences the size of the unit of observation, i.e., the county. Areas with a high population density are split into several counties whereas regions with a low population density are characterized by counties that encompass a vast area. As a consequence, the size of a county is endogenous.<sup>39</sup> To mitigate this issue, we will use dependent variables expressed as per capita ratios or shares which are computed using only the information of the respective county. Nevertheless, the subsequent estimates should be interpreted with caution.

The results in columns (1)–(2) of Table 5 show that our main results are reproducible at the county level: An increase in the local degree of malaria transmission intensity reduces the share of people living in urban areas as well as the share of workers employed in the manufacturing sector in the year 1990. In addition, we find that productivity within the manufacturing sector—measured as the (log) output per worker employed in the sector—is also lower in counties characterized by a high degree of malaria transmission

---

<sup>39</sup>Using regression Eq.(1) to investigate the effect of our malaria transmission suitability on the area covered by a county, we find that a one-standard deviation increase in the MSM coefficient increases the size of a county by 28 percent (evaluated at the average size of the counties).

Table 5: Malaria Suitability and Disparities at the County Level (1990)

Dependent Variable	Urban Share	Industry Share	Log Manufacturing Output per Capita	Log Income per Capita	Han Share	Natural Rate of Increase	Child Share	In-Migration Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MSM (SD)	-0.041** (0.020)	-0.091*** (0.017)	-0.250*** (0.059)	-0.292*** (0.097)	-0.232*** (0.042)	0.001 (0.001)	0.037 (0.110)	-0.010* (0.005)
Climate Controls	yes	yes	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.104	0.548	0.160	0.420	0.513	0.372	0.547	0.223
Obs.	1,148	1,148	958	804	1,147	1,106	1,107	1,105

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, coastline intersection indicator, Pearl and Yangtze river intersection dummies, elevation, rice suitability dummy, wheat suitability dummy, and agricultural suitability. The number of observations vary according to data availability.

intensity (column (3)). These malaria suitability-related differences are also reflected in income per capita (column (4)). The MSM coefficient of  $-0.292$  implies that a one-standard deviation increase in the local malaria potential reduces the county-level income per capita by 25 percent.<sup>40</sup> This result documents that the malaria-influenced spatial pattern of industrialization and urbanization induces substantial regional income disparities.

In column (5), we investigate the ethnic composition of a county's population. The historical difficulties of the Han Chinese in penetrating areas characterized by a high potential for malaria transmission is still reflected in the population structure. An increase in the MSM reduces the share of the total population that belongs to the Han ethnicity.

As a final step, we use demographic information to investigate whether there are any malaria-suitability related differences in population characteristics across counties that could help explain its persistent negative effect on development. In column (6) we regress the natural rate of increase, defined as the difference between the crude birth and death rate in 1990, on our malaria suitability measure. The resulting coefficient estimate is very close to zero and statistically not significant at conventional confidence levels. This indicates that population in regions with a high potential for malaria transmission did not grow faster than in regions with a low malaria potential in 1990. This result captures a snapshot of a single year (1990). The estimate of column (6) therefore does not tell us whether the eradication of malaria led to any differential population growth in the past. To address this question, we compute the size of the cohort aged 0–4 relative to the cohort aged 40–44 for each county. The latter group represents the cohort that was aged 0–4 in the

<sup>40</sup> $e^{-0.292} - 1 = -0.253$

pre-eradication era (i.e., in 1950). Because malaria-related mortality is concentrated among young children, we expect that eradication leads to a shift in the population distribution towards young children, implying that the local malaria suitability is positively associated with the ratio of 0–4 year olds to 40–44 year olds. As unveiled in column (7), there is no statistically significant relationship between the climatic malaria environment and relative cohort sizes.<sup>41</sup> This null-result is likely to be strongly influenced by the implementation of the one-child policy through which the Chinese government directly influenced fertility decisions and population growth dynamics.

A factor that can consolidate the absence of an (expected) disproportionate increase in population in previously malarious areas and the observed persistence in the spatial distribution of urbanization is migration. The result presented in column (8) documents that there is a negative relationship between the malaria transmission potential and share of total population that immigrated during the last five years (i.e., 1985–1990).<sup>42</sup> This indicates that counties with a more favourable malaria environment attract a disproportionate amount of migrants. These counties are, on average, more urbanized and industrialized which suggests the existence of increasing returns to scale. Concentrating population and economic activity is beneficial. The results regarding migration also provide a potential explanation for the persistent effect of the malaria environment on the spatial distribution of urbanization: Historically large urban areas experience a continuous inflow of migrants. This reinforces the existing malaria-influenced spatial pattern of urbanization.

## 7 Conclusion

In this paper, we show that the spatial distribution of economic and urban centers in one of the largest and most rapidly growing and urbanizing economies of the world is influenced by the local malaria environment. We show that the settlement of the Han Chinese in malarious areas was strongly deterred since the start of their colonization of South China. This malaria-influenced spatial pattern of urbanization is still reflected in the today’s distribution. Areas characterized by a high climatic potential for malaria transmission are still relatively unimportant in terms of their urban population and economic activity even though *Plasmodium falciparum* malaria has been largely eradicated since the 1970s.

---

<sup>41</sup>This null result is robust to varying the age groups included in the computation of the ratio.

<sup>42</sup>The migration numbers are directly drawn from official census data and therefore reflect the number of persons that migrate within the hukou system. Temporary migration, for example, is not captured in this metric.

## References

- Abramson, Daniel Benjamin (2006) ‘Urban planning in China: Continuity and change: What the future holds may surprise you.’ *Journal of the American Planning Association* 72(2), 197–215
- Acemoglu, Daron, and Simon Johnson (2007) ‘Disease and development: The effect of life expectancy on economic growth.’ *Journal of Political Economy*, *University of Chicago Press* 115(6), 925–985
- Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001) ‘The colonial origins of comparative development: An empirical investigation.’ *American Economic Review* 91(5), 1369–1401
- Acemoglu, Daron, Simon Johnson, and James Robinson (2003) ‘Disease and development in historical perspective.’ *Journal of the European Economic Association* 1(2-3), 397–405
- Aghion, Philippe, Peter Howitt, and Fabrice Murtin (2011) ‘The Relationship Between Health and Growth: When Lucas Meets Nelson-Phelps.’ *Review of Economics and Institutions*
- Alsan, Marcella (2015) ‘The effect of the TseTse fly on African development.’ *American Economic Review* 105(1), 382–410
- Ashraf, Quamrul H., Ashley Lester, and David N. Weil (2009) *When Does Improving Health Raise GDP?*
- Barrios, Salvador, Luisito Bertinelli, and Eric Strobl (2010) ‘Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy.’ *Review of Economics and Statistics* 92(2), 350–366
- Bleakley, Hoyt (2007) ‘Disease and development: Evidence from Hookworm eradication in the American South.’ *The Quarterly Journal of Economics* 122(1), 73–117
- Bleakley, Hoyt (2010) ‘Malaria eradication in the Americas: A retrospective analysis of childhood exposure.’ *American Economic Journal: Applied Economics* 2(2), 1–45
- Bleakley, Hoyt, and Jeffrey Lin (2012) ‘Portage and path dependence.’ *The Quarterly Journal of Economics* 127(2), 587–644
- Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm (2007) ‘Looking for multiple equilibria when geography matters: German city growth and the WWII shock.’ *Journal of Urban Economics* 61(1), 152 – 169
- Bosker, Maarten, Steven Brakman, Harry Garretsen, and Marc Schramm (2012) ‘Relaxing Hukou: Increased labor mobility and China’s economic geography.’ *Journal of Urban Economics* 72(2), 252–266

- Brinkhoff, Thomas (2015) ‘City population.’ [www.citypopulation.de](http://www.citypopulation.de)
- Buchanan, Keith (1970) *The Transformation of the Chinese Earth* (G. Bell and Sons, LTD)
- Carter, Richard, and Kamini N. Mendis (2002) ‘Evolutionary and historical aspects of the burden of malaria.’ *Clinical Microbiology Reviews* 15(4), 564–594
- Cervellati, Matteo, and Uwe Sunde (2013) ‘Life expectancy, schooling, and lifetime labor supply: Theory and evidence revisited.’ *Econometrica* 81(5), 2055–2086
- Chakraborty, Shankha, Chris Papageorgiou, and Fidel Pérez Sebastián (2010) ‘Diseases, infection dynamics, and development.’ *Journal of Monetary Economics* 57(7), 859–872
- Chan, Kam Wing (2010) ‘Fundamentals of China’s urbanization and policy.’ *The China Review* 10(1), 63–94
- Chan, Kam Wing, J. Vernon Henderson, and Kai Yuen Tsui (2008) ‘Spatial dimensions of Chinese economic development.’ In *China’s Great Economic Transformation*, ed. Loren Brandt and Thomas G. Rawski (Cambridge University Press) pp. 776–828. Cambridge Books Online
- Chang, Sen-Dou (1961) ‘Some aspects of the urban geography of the Chinese hsien capital.’ *Annals of the Association of American Geographers* 51(1), 23–45
- Chang, Sen-Dou (1963) ‘The historical trend of Chinese urbanization.’ *Annals of the Association of American Geographers* 53(2), 109–143
- CHGIS (2010) *Fairbank Center for Chinese Studies and the Institute for Chinese Historical Geography at Fudan University*, Version: 5 ed.
- Chow, Ching-Yen (1991) ‘Malaria vectors in China.’ *Chinese Journal of Entomology* 6, 67–79
- Conley, T. G. (1999) ‘GMM estimation with cross sectional dependence.’ *Journal of Econometrics* 92(1), 1–45
- Davis, Donald R., and David E. Weinstein (2002) ‘Bones, bombs, and break points: The geography of economic activity.’ *American Economic Review* 92(5), 1269–1289
- de Vries, Jan (1976) *The Economy of Europe in an Age of Crisis, 1600-1750* (Cambridge, UK: Cambridge University Press)
- Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (2012) ‘Temperature shocks and economic growth: Evidence from the last half century.’ *American Economic Journal: Macroeconomics* 4(3), 66–95



- Depetris-Chauvin, Emilio, and David N. Weil (2013) ‘Malaria and early African development: Evidence from the sickle cell trait.’ Working Paper 19603, National Bureau of Economic Research, October
- Easterly, William, and Ross Levine (2012) ‘The European origins of economic development.’ Working Paper 18162, National Bureau of Economic Research, June
- Faber, Benjamin (2014) ‘Trade integration, market size, and industrialization: Evidence from China’s national trunk highway system.’ *The Review of Economic Studies* 81(3), 1046–1070
- Flückiger, Matthias, and Markus Ludwig (2016) ‘Malaria suitability, urbanization and regional development in Sub-Saharan Africa.’ Technical Report, Mimeo
- Gallup, John Luke, and Jeffrey D. Sachs (2001) ‘The economic burden of malaria.’ *The American Journal of Tropical Medicine and Hygiene* 64(1 suppl.), 85–96
- Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinger (1999) ‘Geography and economic development.’ *International Regional Science Review* 22(2), 179–232
- Gething, Peter, Thomas Van Boeckel, David Smith, Carlos Guerra, Anand Patil, Robert Snow, and Simon Hay (2011a) ‘Modelling the global constraints of temperature on transmission of Plasmodium falciparum and P. vivax.’ *Parasites & Vectors* 4(1), 92
- Gething, Peter W., Anand P. Patil, David L. Smith, Carlos A. Guerra, Iqbal R.F. Elyazar, Geoffrey L. Johnston, Andrew J. Tatem, and Simon I Hay (2011b) ‘A new world malaria map: Plasmodium falciparum endemicity in 2010.’ *Malaria Journal* 10, 378–378
- Gollin, Douglas, and Christian Zimmermann (2008) ‘Malaria: Disease Impacts and Long-Run Income Differences.’ Center for Development Economics 2008-01, Department of Economics, Williams College, October
- Henderson, J. Vernon (2005) ‘Urbanization and Growth.’ In *Handbook of Economic Growth*, ed. Philippe Aghion and Steven Durlauf, vol. 1 of *Handbook of Economic Growth* (Elsevier) chapter 24, pp. 1543–1591
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil (2012) ‘Measuring economic growth from outer space.’ *American Economic Review* 102(2), 994–1028
- Holcombe, Charles (2001) *Genesis of East Asia, 221 B.C.-A.D. 907* (Asian Interactions and Comparisons)
- Jedwab, Remi, and Alexander Moradi (2016) ‘The permanent effects of transportation revolutions in poor countries: Evidence from Africa.’ *The Review of Economics and Statistics* 98(2), 268–284

- Jedwab, Remi, Edward Kerby, and Alexander Moradi (forthcoming) ‘History, path dependence and development: Evidence from colonial railroads, settlers and cities in Kenya.’ *Economic Journal*
- Jia, Ruixue (2014) ‘The legacies of forced freedom: China’s treaty ports.’ *Review of Economics and Statistics* 96(4), 596–608
- Kalemli-Ozcan, Sebnem (2002) ‘Does the Mortality Decline Promote Economic Growth?’ *Journal of Economic Growth* 7(4), 411–39
- Kalemli-Ozcan, Sebnem, Harl E. Ryder, and David N. Weil (2000) ‘Mortality decline, human capital investment, and economic growth.’ *Journal of Development Economics* 62(1), 1–23
- Killeen, Gerry F., F.Ellis McKenzie, Brian D. Foy, Claus Bøgh, and John C. Beier (2001) ‘The availability of potential hosts as a determinant of feeding behaviours and malaria transmission by African mosquito populations.’ *Transactions of the Royal Society of Tropical Medicine and Hygiene* 95(5), 469 – 476
- Kiple, Kenneth F., ed. (1993) *The Cambridge World History of Human Disease* (Cambridge University Press, Reissue edition)
- Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S. E. Sachs, and J. Sachs (2004) ‘A global index representing the stability of malaria transmission.’ *American Journal of Tropical Medicine and Hygiene* 70(5), 486–498
- Klein Goldewijk, Kees, Arthur Beusen, and Peter Janssen (2010) ‘Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1.’ *The Holocene* 20(4), 565–573
- Lewis, Mark Edward (2011) *China between Empires: The Northern and Southern Dynasties* (Belknap Press)
- Lewis, M.E. (2009) *China’s Cosmopolitan Empire* History of Imperial China (Harvard University Press)
- Lucas, Adrienne M. (2013) ‘The Impact of Malaria Eradication on Fertility.’ *Economic Development and Cultural Change* 61(3), 607 – 631
- Lucas, Robert E. (1988) ‘On the mechanics of economic development.’ *Journal of Monetary Economics* 22(1), 3 – 42
- Ma, Laurence J.C. (2005) ‘Urban administrative restructuring, changing scale relations and local economic development in China.’ *Political Geography* 24(4), 477–497
- Maccini, Sharon, and Dean Yang (2009) ‘Under the weather: Health, schooling, and economic consequences of early-life rainfall.’ *American Economic Review* 99(3), 1006–26

- Manguin, Sylvie, Pierre Carnevale, and Jean Mouchet (2008) *Biodiversity of Malaria in the world* (John Libbey Eurotext)
- Marks, Robert B. (1998) *Tigers, Rice, Silk, and Silt: Environment and Economy in Late Imperial South China* Studies in Environment and History (Cambridge University Press)
- (2011) *China: Its Environment and History* (Rowman & Littlefield Publishers)
- Matteo Cervellati, Uwe Sunde (2005) ‘Human capital formation, life expectancy, and the process of development.’ *The American Economic Review* 95(5), 1653–1672
- McNeill, William H. (1976) *Plagues and Peoples* (Anchor)
- Meng, Xin, Nancy Qian, and Pierre Yared (2015) ‘The institutional causes of china’s great famine, 1959–1961.’ *The Review of Economic Studies*
- Michaels, Guy, and Ferdinand Rauch (forthcoming) ‘Resetting the Urban Network: 117–2012.’ *The Economic Journal*
- Miguel, Edward, and Gerard Roland (2011) ‘The long-run impact of bombing vietnam.’ *Journal of Development Economics* 96(1), 1 – 15
- Mordecai, Erin A., Krijn P. Paaijmans, Leah R. Johnson, Christian Balzer, Tal Ben-Horin, Emily de Moor, Amy McNally, Samraat Pawar, Sadie J. Ryan, Thomas C. Smith, and Kevin D. Lafferty (2013) ‘Optimal temperature for malaria transmission is dramatically lower than previously predicted.’ *Ecology Letters* 16(1), 22–30
- Motamed, Mesbah J., Raymond J.G.M. Florax, and William A. Masters (2014) ‘Agriculture, transportation and the timing of urbanization: Global analysis at the grid cell level.’ *Journal of Economic Growth* 19(3), 339–368
- Ngwa, Gideon A. (2006) ‘On the population dynamics of the malaria vector.’ *Bulletin of Mathematical Biology* 68(8), 2161–2189
- Nordhaus, William D. (2006) ‘Geography and macroeconomics: New data and new findings.’ *Proceedings of the National Academy of Sciences of the United States of America* 103(10), 3510–3517
- Nunn, Nathan (2009) ‘The Importance of History for Economic Development.’ *Annual Review of Economics* 1(1), 65–92
- Nunn, Nathan, and Diego Puga (2012) ‘Ruggedness: The blessing of bad geography in Africa.’ *The Review of Economics and Statistics* 94(1), 20–36
- Nunn, Nathan, and Leonard Wantchekon (2011) ‘The slave trade and the origins of mistrust in Africa.’ *American Economic Review, American Economic Association* 101(7), 3221–52

- Paul Bairoch, J.B.C.P. (1988) *Population des villes européennes de 800 à 1850: banque de données et analyse sommaire des résultats (la)* (Centre d'histoire économique Internationale de l'Université de Genève, Librairie Droz)
- Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney (2002) 'The global distribution of cultivable lands: Current patterns and sensitivity to possible climate change.' *Global Ecology and Biogeography* 11(5), 377–392
- Rappaport, Jordan, and Jeffrey D. Sachs (2003) 'The United States as a coastal nation.' *Journal of Economic Growth* 8(1), 5–46
- Redding, Stephen J., Daniel M. Sturm, and Nikolaus Wolf (2010) 'History and industry location: Evidence from German airports.' *Review of Economics and Statistics* 93(3), 814–831
- Rubel, Franz, and Markus Kottek (2010) 'Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification.' *Meteorologische Zeitschrift* 12(2), 135–141
- Shu, Wei Sheng (1932) 'Medical and health reports. a preliminary survey of the endemicity of malaria in Nanking, Soochow, Hangchow, Wukong and districts along the Yangtze Valley.' *Chinese Medical Journal* 46(7), 719–748
- Simmons, James Stevens, Tom F. Whayne, and Gaylord West Anderson (1944) *India and the Far East: A Geography of Disease and Sanitation* (William Heineman, LTD)
- Skinner, G. W., and Mark Henderson (2012) 'China-A dataset, county-level units 1990'
- Skinner, G. W., Zumou Yue, and Mark Henderson (2011) 'ChinaW—cities, county seats and yamen units (1820 - 1893)'
- (2013) 'China-T (1990) central place analysis'
- Skinner, G. William (1977a) 'Cities and the hierarchy of local system.' In *Studies in Chinese Society*, ed. G. William Skinner (Stanford, California: Stanford University Press) pp. 272–352
- (1977b) 'Regional urbanization in nineteenth-century China.' In *Studies in Chinese Society*, ed. G. William Skinner (Stanford, California: Stanford University Press) pp. 211–252
- Smith, David L., and Nick Ruktanonchai (2010) 'Progress in modelling malaria transmission.' In *Modelling Parasite Transmission and Control*, ed. Edwin Michael and Robert C. Spear, vol. 673 of *Advances in Experimental Medicine and Biology* (Springer New York) pp. 1–12
- Soares, Rodrigo R. (2005) 'Mortality reductions, educational attainment, and fertility choice.' *American Economic Review* 95(3), 580–601

- Spolaore, Enrico, and Romain Wacziarg (2013) ‘How Deep Are the Roots of Economic Development?’ *Journal of Economic Literature* 51(2), 325–69
- Tang, Lin-hua, Hui-lin Qian, and Shu-hui Xu (1992) ‘Malaria and its control in the People’s Republic of China.’ *Southeast Asian Journal of Tropical Medicine and Public Health*
- Tang, Linhua (2000) ‘Progress in malaria control in China.’ *Chinese Medical Journal* 113(1), 89–92
- Vörösmarty, C. J., B. M. Fekete, M. Meybeck, and R. B. Lammers (2000) ‘Global system of rivers: Its role in organizing continental land mass and defining land-to-ocean linkages.’ *Global Biogeochemical Cycles* 14(2), 599–621
- Webb, James L.A. (2008) *Humanity’s Burden: A Global History of Malaria* (Cambridge University Press)
- Weil, David N. (2014) ‘Health and Economic Growth.’ In ‘Handbook of Economic Growth,’ vol. 2 of *Handbook of Economic Growth* (Elsevier) chapter 3, pp. 623–682
- Wiens, Herold J. (1967) *Han Chinese Expansion in South China* (Shoe String Press, Inc., 2ed)
- Yang, Mu, and Hong Yu (2011) *China’s Industrial Development in the 21st Century* Series on contemporary China (World Scientific)
- Young, Alwyn (2005) ‘The gift of the dying: The tragedy of AIDS and the welfare of future African generations.’ *The Quarterly Journal of Economics* 120(2), 423–466
- Zhu, Jian-Ping (2011) ‘Exploration of the relationship between geographical environment and human diseases in ancient China.’ *Journal of Traditional Chinese Medicine* 31(4), 382–385

# Appendices

## A Supporting Material

### A.1 The Chinese Migration Waves

Even though a large part of South China, i.e., the region of today’s China that lies below the Huai River–Qin Mountains line, was conquered by the Chinese during the early Han

era, the population density was relatively low. For example, in the year 2 CE only 20 percent of the Chinese lived in South China; the vast majority were still settled in the northern plains. This started to change with the first major wave of southward migration into the Yangtze valley at the end of the Han dynasty (220 CE) which was triggered by floods and famines as well as wars that took place in northern China (Wen et al., 2004). The southward pressure culminated with the complete destruction of the empire’s capital by nomadic Tatars from Central Asia in 331 CE (Marks (2011, p.106ff.,p.111f.), Wiens (1967, p.175ff.)).

The second major migration wave was set off by the An Lushan as well as the subsequent Wang Hsien-chih and Huang Chao rebellions during the Tang dynasty (618 CE–907 CE) (Hartwell, 1982; Chang, 1963; Wen et al., 2004). Compared to the first major push southward, this second wave was substantially larger (Wen et al., 2004). By 742 CE, the majority of Chinese people (57 percent) lived in South China (Hartwell, 1982). This increase in population density which was accompanied by the establishment of the administrative system in newly settled areas lead to a marked increase in the number of cities (administrative capitals) located within South China (Hartwell, 1982; Chang, 1963) .

By far the largest migration wave into South China was caused by the Mongol invasion in the north of China in the 13th century (Buchanan (1970, p.18-19), Hartwell (1982), Wen et al. (2004)). This wave pushed approximately five million people from North China into the South. After the end of the migration wave, 71 percent of the Chinese population was located south of the Huai River–Qin Mountains line, documenting that the economic center as well as the population mass had shifted from the North into the South over the course of a millennium (Hartwell, 1982). This pattern persists until today.<sup>43</sup>

## A.2 Climate-Based Malaria Transmission Suitability Model

This section provides a detailed description of the construction of the measure of the climatic *Plasmodium falciparum* transmission intensity.<sup>44</sup> We also show that the climatic potential for malaria transmission remains relatively stable over time.

Due to their respective thermal physiologies, air temperature influences both the mosquito’s life cycle and the malaria parasite’s development rate. Air temperature, consequently,

---

<sup>43</sup>For example, in the year 2010 the share of the total urban population living in South China was 58 percent (own calculations based on Brinkhoff (2015)).

<sup>44</sup>This section is directly adopted (with slight modifications) from our recent paper (Flückiger and Ludwig, 2016).

constitutes the most important climatic factor that determines the potential for malaria transmission. Therefore, the spatial extent as well as the intensity of malaria transmission are crucially influenced by temperature.<sup>45</sup> Transmission is generally only possible within a certain temperature range. Within this interval, an optimal temperature exists for the development of the vector and the parasite (Mordecai et al., 2013). Divergence from the optimal temperature results in a slower development of the parasite within the mosquito and also reduces the life expectancy of the mosquito itself. Both of these effects decrease the probability that a mosquito will become infectious. In this study, we draw heavily on Mordecai et al. (2013) to develop a measure that captures the climatic potential for malaria transmission. This measure is based on a temperature-sensitive version of the widely used Ross-MacDonald model.<sup>46</sup> The only factor generating variation in the model is air temperature. The model is fitted to laboratory data on the temperature response of the parasite as well as the mosquito at various development stages.

The metric of the model presented in Mordecai et al. (2013) is the basic reproductive number  $R_0$ , which represents the number of cases of a disease that arise from one case being introduced into a population of susceptible hosts.  $R_0$ , therefore, is a measure for the malaria transmission intensity (Koella, 1991). It is defined as:

$$R_0 = \left( \frac{M a^2 b c e^{-\mu E}}{N r \mu} \right)^{1/2}. \quad (2)$$

This metric consists of three parts. On the one hand, there are (i), components that only characterize aspects of the *Plasmodium falciparum* parasite development and the *Anopheles* mosquito physiology. These are: the mosquito density ( $M$ ), the mosquito mortality rate ( $\mu$ ), and the extrinsic incubation period of the *Plasmodium falciparum* parasite in the mosquito ( $E$ ). On the other hand, there are (ii), characteristics of the *Anopheles* that indirectly depend on attributes of the human population. These factors are the human biting rate of the *Anopheles* ( $a$ ), the proportion of the bites of infective mosquitoes that infect susceptible humans ( $b$ ), and the proportion of bites of susceptible mosquitoes on infectious humans that infect mosquitoes ( $c$ ). Further, there are (iii) components that depend only on human physiology, specifically, human population density ( $N$ ), and the human recovery rate ( $r$ ). To ensure that our measure is not endogenous with respect to

---

<sup>45</sup>See, e.g., Gething et al. (2011) or Mordecai et al. (2013) for a detailed discussion regarding the crucial role of air temperature in determining the spatial extent and local intensity of malaria *falciparum* transmission.

<sup>46</sup>For an overview of this type of mathematical model, see Koella (1991).

the human population, we hold  $N$  and  $r$  constant ( $N = r = 1$ ). Since the vector characteristics of components (ii) are closely related to the characteristics of the human population, we deviate from Mordecai et al. (2013) and only parametrize the components of (i) as a function of air temperature,  $T$ . That is, we also set  $a = b = c = 1$ .<sup>47</sup>

With the above-mentioned modifications, Eq.(2), now written as a temperature-dependent function, becomes:

$$R_0(T) = \left( \frac{M(T) e^{-\mu(T) E(T)}}{\mu(T)} \right)^{1/2} = \left( \frac{e^{-\frac{\mu(T)}{PDR(T)}} EFD(T) p_{EA}(T) MDR(T)}{\mu^3(T)} \right)^{1/2}. \quad (3)$$

In the second term of the right-hand side, we express  $M(T)$ , the mosquito density, as a function of the temperature-dependent vector development (Mordecai et al., 2013, p. 25).<sup>48</sup> This substitution implies that the measure is now fully characterized by the thermal physiology of the vector and the parasite.  $PDR(T)$  represents the *Plasmodium falciparum* development rate,<sup>49</sup>  $EFD(T)$  denotes the number of eggs laid per female mosquito,  $p_{EA}(T)$  the probability that a mosquito egg develops into an adult mosquito, and,  $MDR(T)$  the larval mosquito development rate. Each of these components represents a non-monotonic function of air temperature. Of critical importance for our study is the fact that the parametrization and the functional forms of these individual components are based solely on experimental data from laboratory studies that analyze the thermal physiology of the parasite and the vector. The results of these studies are—unlike values derived from field studies—independent of local differences in human population densities. The functional forms and parameter values derived in the laboratory studies are gathered in Mordecai et al. (2013).<sup>50</sup> We adopt them directly.

Figure A.1, Panel (a) depicts the relationship between temperature and malaria transmission intensity. The optimal temperature for malaria transmission lies around 25 degrees Celsius. At temperatures below 17°C and above 34°C, the climatic potential for malaria transmission is zero.

Our goal is to analyze the effect of climatic malaria transmission suitability on the geo-

---

<sup>47</sup>Mordecai et al. (2013) define  $b$ ,  $c$  and  $a$  to be temperature-sensitive due to the fact that these components—apart from capturing characteristics of the human population—are also influenced by the anopheles vector physiology (which is temperature-sensitive).

<sup>48</sup>More specifically, the mosquito density is expressed as  $M(T) = \frac{EFD(T) p_{EA}}{\tau_{EA}(T) \mu(T)^2}$ ; where  $\tau_{EA}(T) = \frac{1}{MDR(T)}$ .

<sup>49</sup>Where  $PDR(T) = \frac{1}{E(T)}$ .

<sup>50</sup>The functions together with their parameter values are depicted in Table 2 in Mordecai et al. (2013).



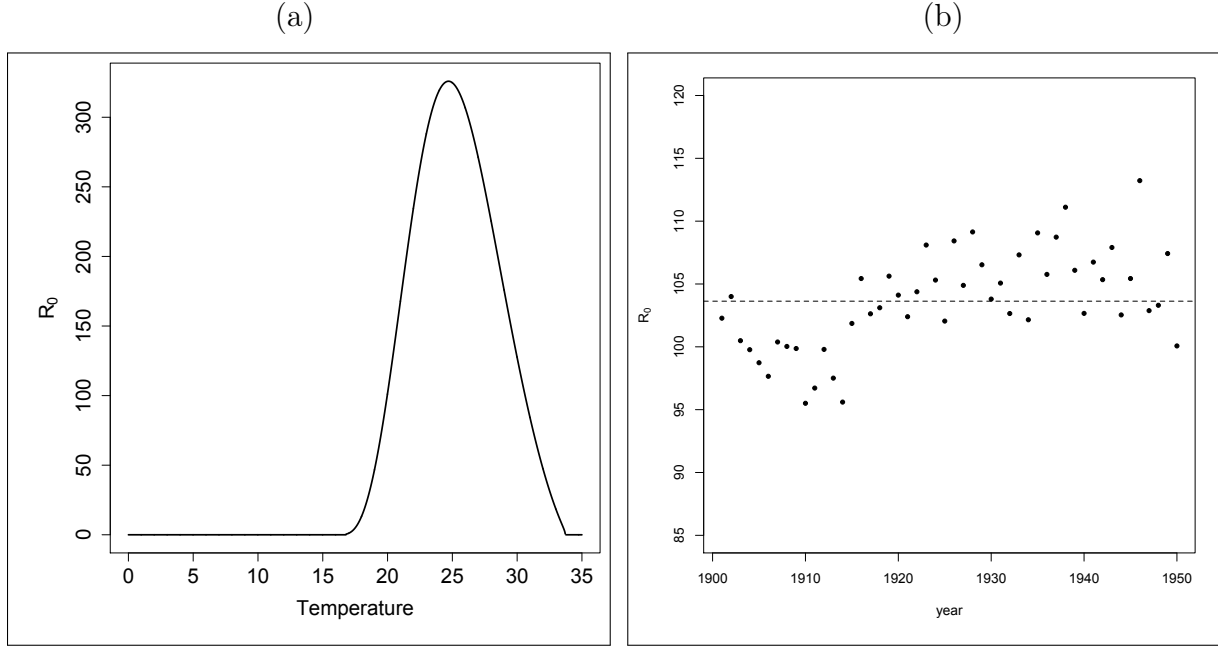


Figure A.1: Panel (a) represents  $R_0$  (Eq.(3)) as a function of air temperature. Panel (b) depicts the (mean) value of  $R_0$  in South China over the time period 1901–1950.

graphical distribution of political, economic and urban centers in China. Ideally, we would want to construct this measure using climate data that reaches back to the start of the unified Chinese empire. However, high-resolution climate data are only available from 1901 onwards. This could raise concerns that potential climate fluctuations—not captured in our data—might drive our results. At the end of this section, we provide evidence that this is not likely to be the case.

We compute the (time-invariant) malaria potential by taking the mean of the monthly malaria transmission intensity,  $R_0$  defined in Eq.(3), over the time period 1901–1950, i.e., the pre-eradication period for which the climate data are available. Formally, this measure can be expressed as:

$$MSM = \frac{1}{600} \sum_{y=1901}^{Y=1950} \sum_{m=1}^{M=12} R_{0,m,y}. \quad (4)$$

$MSM$  represents our measure for the average malaria transmission suitability. It is computed as the sum of the monthly malaria transmission intensity  $R_{0,m,y}$  in month  $m$  of year  $y$  over the time span 1901–1950. This value is then divided by the number of months in this time period, which is 600. In the empirical analysis, we will use the standardized  $MSM$ , i.e., the z-score, as the independent variable.

Panel (b) of Figure A.1 shows that the (average) degree of malaria transmission suitability in South China is relatively stable over time. This gives a first indication that our results are not driven by the choice of the time span used to construct our measure. This is further supported by the fact that our results remain stable to varying the time span (Tables C.13–C.16). Using the data of Mann et al. (2009) which is aggregated at the  $5\times 5$  degree grid-cell-level and spans the time-period 500 CE–2006 CE, we can further show that the relative position of a grid cell in the distribution of the suitability measure remains stable over time. To this end, we first compute the relative position in the malaria suitability distribution of each  $5\times 5$  degree cell for each year by dividing its value by the average suitability in the corresponding year. We then separately regress the MSM measure constructed from temperature data in the interval 1901–1950 on the MSM-values constructed using the climate data for the years 500–1800 (at an interval of 100 years). These regressions are depicted in Figure A.2. As can be seen, the coefficients are close to one for all years, indicating that a grid cell’s relative position in the MSM distribution is representative for all time periods (between 500 and 1950). Further supporting our assumption, we find that the serial correlation in the MSM measure is 0.974 (employing the  $0.5\times 0.5$  degree data) and 0.992 (employing the  $5\times 5$  degree data).

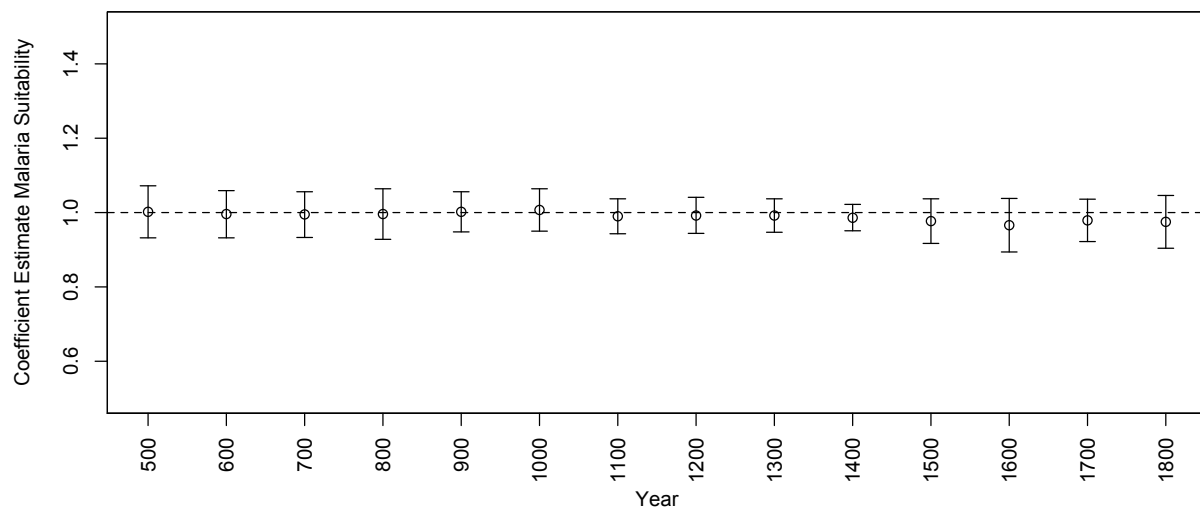


Figure A.2: Graphic representation of the relationship between the MSM measure constructed using current temperature (1901–1950) values and the MSM measure constructed using past temperature values. The number of observations for each regression is 28.

## B Assigning County-Level Industry Information to Grid Cells

This section describes the procedure employed to assign the county-level industry employment provided in the Skinner et al. (2011) dataset to the  $0.5 \times 0.5$  degree grid cells. Whenever a county's territory lies entirely within a single grid cell, we simply assign the county-level information, i.e., the number of workers in the industry sector, to the corresponding cell. If a county overlaps multiple grid cells, the values are assigned to the cells in proportion to the share of the county's total population that lives within the respective grid cell. The location as well as the size of the settlements within each county are extracted from Skinner et al. (2013). The fact that the population numbers in the Skinner et al. (2011) and the Skinner et al. (2013) dataset are derived from the same census data ensures the compatibility of the two data sources. The assumption underlying the assignment procedure is that industrial activity is proportional to the population size of the settlements.

We illustrate the assignment procedure using Figure B.1 which depicts the area encompassed by four  $0.5 \times 0.5$  degree grid cells. The grid-cell boundaries are represented by the dashed lines. The boundaries of the counties are drawn as solid lines. As can be seen, each county intersects multiple grid cells. The industry-related county-level information is assigned to the intersected grid cells according to their share of the county's total population. The locations of the settlements are represented as dots, where their population numbers are reflected in the size of the dots. The shading represents the share of the county-level value assigned to each grid cell.

To further clarify the procedure, take county A. This county overlaps all grid cells, i.e., grid cells 1–4. However, the county is only home to three populated places which are spread over two grid cells. The two larger places lie in grid cell 1, whereas the smaller one is located in cell 3. The weight for grid cell 1 is computed by dividing the number of people living in cell 1 and county A by the total population of county A. The weight for grid cell 3 is computed analogously. Grid cells 2 and 4 both receive weights of zero due to the fact that county A has no populated place located in these cells. Using these weights, we then compute the grid-cell-specific value of industrial employment for each county by multiplying the county-level value with the grid-cell specific weights. The result of the assignment procedure just described is a dataset at the county-grid-cell level. We then aggregate the data to the grid-cell level by computing the sum.

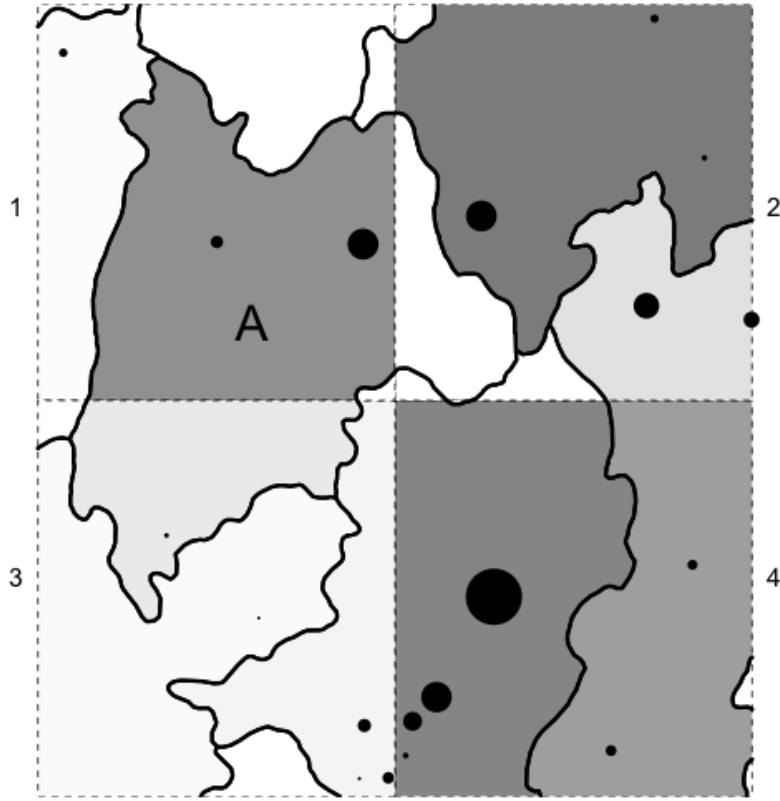


Figure B.1: Assignment Procedure County to Grid Cell

Table B.1: Descriptive Statistics County-Level Variables

Variable	Mean	Std. Dev.	Min.	Max.	Obs.
Urban Share	0.152	0.154	0.016	0.953	1148
Industry Share	0.459	0.117	0.176	0.784	1148
Log Manufacturing Output per Capita	0.822	0.475	-1.984	3.221	958
Log Income per Capita	-2.367	0.587	-4.919	0.267	804
Han Share	0.826	0.290	0.005	1.000	1147
Natural Rate of Increase	0.021	0.007	-0.050	0.046	1106
Child Share	2.069	0.657	0.708	4.154	1107
In-Migration Share	0.028	0.045	0.002	0.066	1105
Malaria Suitability (SD)	0	1	-2.965	3.503	1171

## B.1 Urbanization across Time and Space

As described in Section 5, we have much more detailed information regarding the distribution of the urban population for the year 1990 compared to the year 1893. For the latter, we only know the size of the population of the administrative capitals, whereas for the year 1990 we have complete information regarding the size of urban places, which also includes many smaller urban places that are not reported in the 1893 data. Either these places did



Figure B.2: Figure depicts the spatial distribution of *Plasmodium falciparum* malaria prevalence for the pre-eradication era (geocoded from Tang et al. (1992)). The dots represent areas in which prevalence was reported. The gray line indicates the flow of the Yangtze river. Clearly visible is the clustering of malaria cases in the lower Yangtze area, one of China's most densely populated regions.

not exist in 1893 or were not home to administrative seats. However, as expected from the outline in Section 3, the pre-modern urban hierarchy exhibits a high degree of persistence. Cities that were large in 1893 are still large in 1990. This is illustrated in Figure B.3 where we plot the distribution of the (log) population of the urban places in 1990 separately for cities that are already observed in 1893, i.e., the then-administrative capitals, and those that are only reported in the 1990 data. It is clearly visible that the cities that were home to an administrative seat in the year 1893 are larger than cities/places that were not. The first group of cities is, on average, more than 13 times bigger than the latter group. The importance of the large imperial cities is also reflected in the fact that of the nine cities in South China that had more than 512,000 inhabitants in 1893, seven of them are still among the nine largest cities in 1990. The fact that large cities in 1990 were for the most part already large cities in the imperial era suggests that the spatial distribution of the (relative) urban population has remained stable since 1893. This is further supported by Figure B.4.

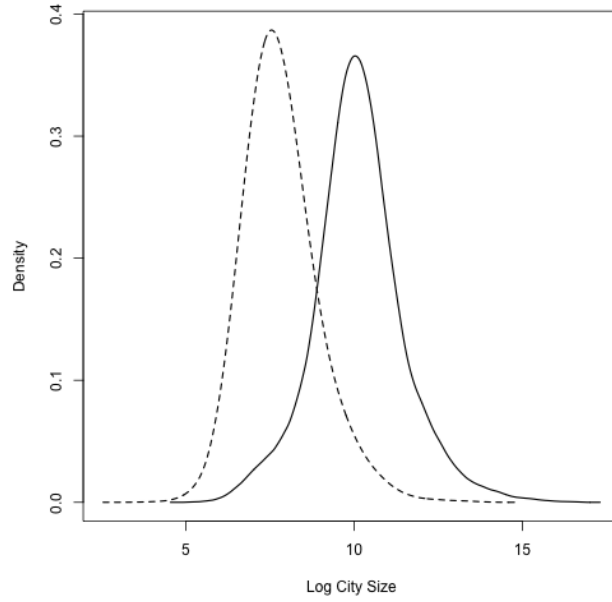


Figure B.3: Density of log city size population in the year 1990 for the urban places that were administrative capitals in 1893 (bold), and urban places that were not (dashed).

Panel (a) and (b) depict the spatial distribution of the urban population in South China for the years 1893 and 1990. The size of the square is representative of the share of total urban population living within a grid cell. Generally speaking, the pattern of relative importance

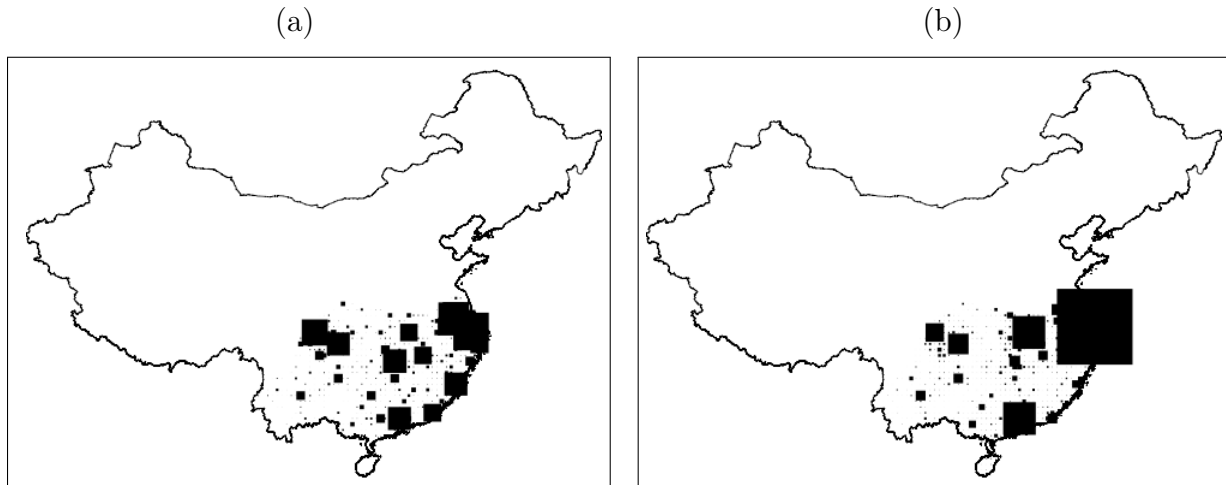


Figure B.4: Panel (a) depicts the spatial distribution of the urban population in the year 1893. Panel (b) depicts it in the year 1990.

is stable. Grid cells that contained a relatively high share of urban population in 1893 still contain a high share today. The two figures suggest a shift in the relative importance

towards the coastal regions. This is consistent with observations of other authors (e.g., Wu and Gaubatz (2012)).<sup>51</sup> A strong degree of persistence in the spatial urban hierarchy is also detectable in a more formal analysis. Regressing the grid cell's share of total urban population in 1990 on the analogous variable computed for the year 1893 in a bivariate setup, we obtain a coefficient of 0.822; the  $R^2$  is 0.534. This high explanatory power is particularly impressive when taking into account that China has undergone a substantial transformation in its political and economic system since 1893, including its transition to becoming an industrial nation. This documents the strong degree of persistence in the spatial distribution of urbanization and suggests that original (historical) factors that lead to the emergence of the spatial urban equilibrium up to 1893 are still reflected in today's geographical distribution of economic activity.

## C Robustness Checks

This section presents various robustness checks which support the general validity of our results.

### C.1 Malaria Suitability Effect over Time

### C.2 South China

---

<sup>51</sup>Our data potentially exaggerates the strength of the eastward shift for two reasons: First, population numbers for the year 1893 are only available as categorical data. This implies that the size of large cities, such as Fuzhou, Guangzhou or Shanghai is capped in the year 1893. Second, the population data represent the number of persons living within the regular Chinese administrative system. Non-Chinese residents in the treaty ports, for example, are not captured in these statistics. Both of these facts potentially lead to an underestimation of the relative importance of the coastal area in 1893. Consequently, the magnitude of the shift in the urban hierarchy to the east between 1893 and 1990 observed in our data constitutes an upper bound.

Table C.1: Malaria Effect 200 BCE – 1900 CE.

Panel A: South China					Panel B: China Proper				Panel C: North China			
Year	Coefficient	Se	R <sup>2</sup>	Obs.	Coefficient	Se	R <sup>2</sup>	Obs.	Coefficient	Se	R <sup>2</sup>	Obs.
-200	-0.258	(0.247)	0.292	868	-0.170	(0.210)	0.406	1489	-0.546	(0.619)	0.463	621
-100	-0.660***	(0.249)	0.215	868	-0.310*	(0.185)	0.433	1489	-0.722	(0.463)	0.525	621
0	-0.694***	(0.239)	0.241	868	-0.259	(0.179)	0.457	1489	-0.665	(0.464)	0.534	621
100	-0.790***	(0.241)	0.227	868	-0.381**	(0.180)	0.417	1489	-0.516	(0.477)	0.511	621
200	-0.925***	(0.221)	0.228	868	-0.511***	(0.180)	0.404	1489	-0.566	(0.487)	0.511	621
300	-0.839***	(0.195)	0.259	868	-0.483***	(0.161)	0.340	1489	-0.396	(0.335)	0.486	621
400	-0.681***	(0.183)	0.256	868	-0.363**	(0.162)	0.297	1489	-0.040	(0.314)	0.451	621
500	-0.566***	(0.186)	0.262	868	-0.186	(0.174)	0.297	1489	0.304	(0.312)	0.439	621
600	-0.390*	(0.213)	0.325	868	-0.067	(0.167)	0.404	1489	-0.099	(0.348)	0.551	621
700	-0.764***	(0.242)	0.240	868	-0.361*	(0.205)	0.289	1489	0.094	(0.263)	0.507	621
800	-0.670***	(0.208)	0.281	868	-0.308*	(0.176)	0.343	1489	0.314	(0.265)	0.545	621
900	-0.654***	(0.208)	0.265	868	-0.270	(0.178)	0.328	1489	0.264	(0.266)	0.540	621
1000	-0.777***	(0.202)	0.272	868	-0.435**	(0.174)	0.353	1489	0.155	(0.286)	0.527	621
1100	-0.775***	(0.215)	0.247	868	-0.494***	(0.181)	0.336	1489	-0.027	(0.290)	0.506	621
1200	-0.688***	(0.212)	0.246	868	-0.432**	(0.182)	0.330	1489	0.036	(0.283)	0.490	621
1300	-0.592***	(0.177)	0.192	868	-0.400**	(0.155)	0.335	1489	-0.106	(0.300)	0.513	621
1400	-0.535***	(0.164)	0.213	868	-0.371**	(0.144)	0.328	1489	-0.121	(0.306)	0.482	621
1500	-0.482***	(0.164)	0.227	868	-0.333**	(0.145)	0.324	1489	-0.110	(0.302)	0.469	621
1600	-0.519***	(0.152)	0.244	868	-0.389***	(0.133)	0.337	1489	-0.138	(0.292)	0.477	621
1700	-0.427***	(0.145)	0.253	868	-0.284**	(0.136)	0.346	1489	-0.083	(0.309)	0.486	621
1800	-0.362**	(0.152)	0.242	868	-0.223*	(0.134)	0.326	1489	0.032	(0.296)	0.463	621
1900	-0.357**	(0.149)	0.246	868	-0.215	(0.133)	0.324	1489	0.028	(0.293)	0.461	621

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at 2×2 degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl, Yangtze and Yellow river intersection dummies, elevation, area of grid cell, and agricultural suitability.



Table C.2: Malaria Effect in South China 200 BCE – 1900 CE: Successively Adding Controls

Year	Panel A: Base Controls <sup>a</sup>				Panel B: Climate Controls <sup>b</sup>				Panel C: Climate and Geography Controls <sup>c</sup>			
	Coefficient	Se	R <sup>2</sup>	Obs.	Coefficient	Se	R <sup>2</sup>	Obs.	Coefficient	Se	R <sup>2</sup>	Obs.
-200	-0.290	(0.204)	0.180	868	-0.236	(0.219)	0.187	868	-0.355	(0.263)	0.230	868
-100	-0.605***	(0.202)	0.095	868	-0.525**	(0.202)	0.115	868	-0.637***	(0.234)	0.154	868
0	-0.657***	(0.195)	0.115	868	-0.566***	(0.195)	0.137	868	-0.686***	(0.233)	0.178	868
100	-0.724***	(0.197)	0.124	868	-0.647***	(0.200)	0.140	868	-0.820***	(0.237)	0.177	868
200	-0.775***	(0.190)	0.137	868	-0.741***	(0.190)	0.141	868	-0.954***	(0.234)	0.170	868
300	-0.640***	(0.162)	0.151	868	-0.645***	(0.163)	0.153	868	-0.863***	(0.205)	0.190	868
400	-0.601***	(0.155)	0.129	868	-0.614***	(0.159)	0.133	868	-0.755***	(0.195)	0.180	868
500	-0.721***	(0.170)	0.142	868	-0.674***	(0.169)	0.155	868	-0.756***	(0.185)	0.206	868
600	-0.437**	(0.178)	0.233	868	-0.399**	(0.176)	0.239	868	-0.441*	(0.226)	0.296	868
700	-0.508**	(0.238)	0.152	868	-0.531**	(0.237)	0.156	868	-0.513**	(0.253)	0.180	868
800	-0.419**	(0.198)	0.184	868	-0.485**	(0.201)	0.199	868	-0.549**	(0.222)	0.230	868
900	-0.395**	(0.196)	0.169	868	-0.456**	(0.199)	0.183	868	-0.525**	(0.224)	0.209	868
1000	-0.541***	(0.173)	0.174	868	-0.616***	(0.171)	0.200	868	-0.591***	(0.222)	0.205	868
1100	-0.569***	(0.180)	0.156	868	-0.642***	(0.176)	0.182	868	-0.630***	(0.225)	0.191	868
1200	-0.503***	(0.179)	0.151	868	-0.584***	(0.174)	0.184	868	-0.548**	(0.222)	0.192	868
1300	-0.520***	(0.142)	0.104	868	-0.599***	(0.143)	0.134	868	-0.533***	(0.169)	0.154	868
1400	-0.490***	(0.130)	0.128	868	-0.573***	(0.129)	0.162	868	-0.516***	(0.162)	0.179	868
1500	-0.445***	(0.129)	0.143	868	-0.515***	(0.133)	0.169	868	-0.431**	(0.163)	0.184	868
1600	-0.503***	(0.130)	0.167	868	-0.569***	(0.127)	0.191	868	-0.490***	(0.151)	0.206	868
1700	-0.439***	(0.121)	0.183	868	-0.494***	(0.121)	0.199	868	-0.442***	(0.134)	0.212	868
1800	-0.349***	(0.124)	0.168	868	-0.411***	(0.121)	0.191	868	-0.351**	(0.144)	0.204	868
1900	-0.337***	(0.120)	0.166	868	-0.398***	(0.118)	0.189	868	-0.341**	(0.143)	0.204	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses.

<sup>a</sup> Base Controls include temperature, temperature squared, longitude, latitude, and area of grid cell.

<sup>b</sup> Climate Controls include precipitation, precipitation squared plus all base controls and the interactions between the base controls as well as the interaction between temperature and precipitation.

<sup>c</sup> Climate and Geography Controls include all climate controls plus distance to coast, Pearl, Yangtze and Yellow river intersection dummies, elevation, and agricultural suitability.

Table C.3: Malaria Suitability Effect on Level of Hierarchy: Ordered Logit

Dependent Variable:	Ordered Logit	OLS
	Hierarchy Level 1893	Urban Population 1990 Klein Goldewijk et al. (2010) Database
	(1)	(2)
MSM (SD)	-1.308*** (0.335)	-1.048*** (0.297)
Climate Controls	yes	yes
Geography Controls	yes	yes
Pseudo R-squared	0.077	
R-squared		0.146
Obs.	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Ordered logit as well as linear regression in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.4: Robustness Controlling for Additional Locational Fundamentals: Wheat Suitability

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.449*** (0.158)	-0.896** (0.352)	-0.329*** (0.111)	-0.702** (0.288)	-0.770*** (0.240)	-1.030** (0.401)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.191	0.164	0.147	0.132	0.288	0.140
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.5: Robustness Controlling for Additional Locational Fundamentals: Rice Suitability

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.522*** (0.168)	-0.776*** (0.291)	-0.406*** (0.112)	-0.592** (0.271)	-0.507** (0.210)	-0.766** (0.330)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.191	0.164	0.142	0.132	0.288	0.139
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.6: Robustness Controlling for Additional Locational Fundamentals: Tea Suitability

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.500*** (0.161)	-0.765** (0.290)	-0.425*** (0.105)	-0.745*** (0.229)	-0.663*** (0.198)	-0.858*** (0.309)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.191	0.166	0.142	0.135	0.295	0.147
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.7: Robustness Controlling for Additional Locational Fundamentals: Soil Type

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.489*** (0.161)	-0.909*** (0.296)	-0.417*** (0.103)	-0.877*** (0.285)	-0.794*** (0.254)	-1.081*** (0.396)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.195	0.188	0.147	0.147	0.303	0.155
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.8: Robustness Controlling for Additional Locational Fundamentals: Geological Regions

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.460*** (0.159)	-0.495** (0.213)	-0.398*** (0.101)	-0.517*** (0.177)	-0.459*** (0.168)	-0.571** (0.252)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.229	0.218	0.184	0.168	0.323	0.168
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.9: Robustness Alternative Malaria Suitability Measure: Gething et al. (2011)

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.564*** (0.141)	-0.687*** (0.248)	-0.502*** (0.093)	-0.549** (0.275)	-0.311 (0.231)	-0.606* (0.322)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.204	0.164	0.160	0.131	0.284	0.139
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.10: Robustness Alternative Malaria Suitability Measure: Mordecai et al. (2013)

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.475** (0.189)	-0.818** (0.348)	-0.439*** (0.125)	-0.824*** (0.297)	-0.776*** (0.269)	-1.029** (0.430)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.182	0.163	0.134	0.131	0.287	0.140
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.11: Robustness Clustering Approaches

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.497 (0.182)** [0.144]***	-0.744 (0.248)*** [0.272]***	-0.428 (0.113)*** [0.096]***	-0.719 (0.237)*** [0.232]***	-0.631 (0.244)** [0.205]***	-0.808 (0.405)* [0.307]***
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.190	0.164	0.141	0.132	0.287	0.139
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at the province level are reported in parentheses. Values in brackets represent standard errors computed using the approach proposed by Conley (1999) (cutoff: 2 degrees). Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.12: Robustness: Temperature Controls Only

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.404*** (0.153)	-1.194*** (0.295)	-0.329*** (0.106)	-1.240*** (0.372)	-1.208*** (0.335)	-1.716*** (0.561)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	no	no	no	no	no	no
R-squared	0.100	0.073	0.081	0.058	0.142	0.062
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions include controls for area of grid cell, absolute latitude, longitude, temperature, and temperature squared. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses.

Table C.13: Robustness Time Interval Climate Data Construction: 1901–1925

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.506*** (0.171)	-0.768** (0.297)	-0.439*** (0.114)	-0.836*** (0.257)	-0.682*** (0.231)	-0.960** (0.372)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.183	0.164	0.138	0.133	0.289	0.139
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.14: Robustness Time Interval Climate Data Construction: 1926–1950

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.473*** (0.149)	-0.808*** (0.273)	-0.382*** (0.098)	-0.763*** (0.226)	-0.651*** (0.198)	-0.793** (0.305)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.190	0.165	0.139	0.133	0.287	0.145
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.15: Robustness Time Interval Climate Data Construction: 1951–1975

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.521*** (0.163)	-0.857*** (0.290)	-0.433*** (0.109)	-0.903*** (0.239)	-0.748*** (0.214)	-1.001*** (0.350)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.186	0.164	0.140	0.134	0.284	0.142
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

Table C.16: Robustness Time Interval Climate Data Construction: 1976–2000

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.519*** (0.157)	-0.739** (0.282)	-0.432*** (0.110)	-0.814*** (0.244)	-0.668*** (0.222)	-0.994*** (0.371)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.183	0.161	0.138	0.133	0.283	0.141
Obs.	868	868	868	868	868	868

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Pearl and Yangtze river intersection dummies, elevation, area of grid cell, and agricultural suitability.

### C.3 North China

The results in this Section document that our climatic malaria transmission suitability measure has no effect on the spatial distribution of the administrative seats, urban population or manufacturing employment. This is consistent with the fact that there is no vector permanently present in North China capable of (efficiently) transmitting *Plasmodium falciparum* malaria (see Section 3). Consequently, the MSM does represent a negative locational fundamental that deters settlement.

Table C.17: Malaria Suitability Effects in North China

Dependent Variable	Number of Seats 1893	Urban Population 1893	Hierarchy Level 1893	Urban Population 1990	Manufacturing Employment 1990	Urban Population 2010
	(1)	(2)	(3)	(4)	(5)	(6)
MSM (SD)	-0.108 (0.237)	-0.324 (0.470)	-0.062 (0.122)	-0.032 (0.450)	0.276 (0.365)	-0.079 (0.455)
Climate Controls	yes	yes	yes	yes	yes	yes
Geography Controls	yes	yes	yes	yes	yes	yes
R-squared	0.499	0.134	0.336	0.092	0.145	0.093
Obs.	621	621	621	621	621	621

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Regressions in all columns with Macroregion, river basin and climate zone dummies. Standard errors clustered at  $2 \times 2$  degree grid cells reported in parentheses. Climate controls include precipitation, precipitation squared, temperature, temperature squared, and precipitation and temperature interaction. Geography controls include distance to coast, Yangtze and Yellow river intersection dummies, elevation, area of grid cell, and agricultural suitability.



## References

- Brinkhoff, Thomas (2015) ‘City population.’ [www.citypopulation.de](http://www.citypopulation.de)
- Buchanan, Keith (1970) *The Transformation of the Chinese Earth* (G. Bell and Sons, LTD)
- Chang, Sen-Dou (1963) ‘The historical trend of Chinese urbanization.’ *Annals of the Association of American Geographers* 53(2), 109–143
- Conley, T. G. (1999) ‘GMM estimation with cross sectional dependence.’ *Journal of Econometrics* 92(1), 1–45
- Flückiger, Matthias, and Markus Ludwig (2016) ‘Malaria suitability, urbanization and regional development in Sub-Saharan Africa.’ Technical Report, Mimeo
- Gething, Peter, Thomas Van Boeckel, David Smith, Carlos Guerra, Anand Patil, Robert Snow, and Simon Hay (2011) ‘Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*.’ *Parasites & Vectors* 4(1), 92
- Hartwell, Robert M. (1982) ‘Demographic, political, and social transformations of China.’ *Harvard Journal of Asiatic Studies* 42(2), 365–442
- Klein Goldewijk, Kees, Arthur Beusen, and Peter Janssen (2010) ‘Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1.’ *The Holocene* 20(4), 565–573
- Koella, Jacob C. (1991) ‘On the use of mathematical models of malaria transmission.’ *Acta Tropica* 49(1), 1–25
- Mann, Michael E., Zhihua Zhang, Scott Rutherford, Raymond S. Bradley, Malcolm K. Hughes, Drew Shindell, Caspar Ammann, Greg Faluvegi, and Fenbiao Ni (2009) ‘Global signatures and dynamical origins of the little ice age and medieval climate anomaly.’ *Science* 326(5957), 1256–1260
- Marks, Robert B. (2011) *China: Its Environment and History* (Rowman & Littlefield Publishers)
- Mordecai, Erin A., Krijn P. Paaijmans, Leah R. Johnson, Christian Balzer, Tal Ben-Horin, Emily de Moor, Amy McNally, Samraat Pawar, Sadie J. Ryan, Thomas C. Smith, and Kevin D. Lafferty (2013) ‘Optimal temperature for malaria transmission is dramatically lower than previously predicted.’ *Ecology Letters* 16(1), 22–30
- Skinner, G. W., Zumou Yue, and Mark Henderson (2011) ‘ChinaW—cities, county seats and yamen units (1820 - 1893)’
- (2013) ‘China-T (1990) central place analysis’

- Tang, Lin-hua, Hui-lin Qian, and Shu-hui Xu (1992) 'Malaria and its control in the People's Republic of China.' *Southeast Asian Journal of Tropical Medicine and Public Health*
- Wen, Bo, Hui Li, Daru Lu, Xiufeng Song, Feng Zhang, Yungang He, Feng Li, Yang Gao, Xianyun Mao, Liang Zhang, Ji Qian, Jingze Tan, Jianzhong Jin, Wei Huang, Ranjan Deka, Bing Su, Ranajit Chakraborty, and Li Jin (2004) 'Genetic evidence supports demic diffusion of Han culture.' *Nature* 431(7006), 302–305
- Wiens, Herold J. (1967) *Han Chinese Expansion in South China* (Shoe String Press, Inc., 2ed)
- Wu, Weiping, and Piper Rae Gaubatz (2012) *The Chinese City* (Routledge)